PhD Dissertation Synopsis

ESSAYS IN MARKET MICROSTRUCTURE

ON THE INFORMATION CONTENT OF QUOTES AND CONTINUOUS AUCTIONS IN ORDER DRIVEN MARKETS

Sudhanshu Sekhar Pani

Doctor of Philosophy (Management)

Reg. No. 80121160008 (Date: 28th November 2019)

School of Business Management, NMIMS, Mumbai – 400 056

Theses Advisory Committee

Dr. Ranjan R Chakravarty Sr. Professor, SBM, NMIMS	Chairperson
Dr. Ramesh Bhat Vice Chancellor, NMIMS	Member
Dr. Chandan Dasgupta Sr. Professor, SBM, NMIMS	Member
Dr. Mayank Joshipura Professor, SBM, NMIMS	Member

CONTENTS

Chapter 1 2
Introduction 2
Chapter 2 11
Intraday trading in order driven markets 11
The Model 16
Chapter 3 22
A Theory of 'Auction as a Search' in Speculative Markets 22
Chapter 4 31
Liquidity in High Resolution in Limit Order Markets 31
The Model 33
Results and Discussions 37
Chapter 5 46
A Data paradigm to handle high resolution Orders and Trades data 46
Chapter 6 49
Conclusion, Limitations and Future Research Opportunities 49
References 52

CHAPTER 1

INTRODUCTION

In the twenty-first century, the organisation of trading in financial markets has changed due to electronic limit order markets. These markets offer all participants a real-time glimpse of the existing supply and demand, a piece of information that used to be earlier available with the dealers and specialist market makers only. While the advent of new technologies of information and communication technology has been the key to enable trading automation, regulatory actions enforcing transparency in trading and competition among exchanges (Reg NMS in the U.S,

MiFID in the E.U) have fostered its evolution. New trading platforms continue to emerge in the business of exchanges leading to competition for order flow in mature markets and fragmentation of markets.

Low latency trading and exchange infrastructure are among the foundation of modern markets. The technology-led transformation of the financial markets continues unabated.

Understanding the consequences of the recent changes in financial markets organisation is a primary interest for academic research. Researchers investigated a few such evolutions in the last two decades, such as algorithmic trading, high frequency trading, fragmented markets, dark pools, and inverted markets. It is evident from the literature that market developments have consistently predated theoretical and empirical research in understanding the mechanisms involved and the analytical tools required to understand the various elements that constitute the workings of such markets.

Data available from the order-driven markets today can help us investigate the markets in high resolution (sub-second time scale). The study of the dynamics of markets in high resolution is an emerging trend; that exploits the availability of high resolution 'trades and quotes' data in datasets provided by the exchanges. A consequence is an inclusion in the research design of information from the limit orders (order placement and cancellation). O'Hara (2015) called for the market microstructure toolbox to be enhanced to investigate 'orders' that were the unit of information in today's markets.

Several recent papers implement objectives that require a high-resolution analysis, including modelling imperfect competition and quote cancellation (Baruch and Glosten (2019)), price discovery in high resolution (Hasbrouck (2019)), price discovery in quotes (Brogaard et al. (2019)), trading in the absence of designated market maker (Kyle et al. (2018)).

The dataset (refer figure 1 and figure 2) from modern markets consists of two crucial additions: First, the timestamp available has a resolution of microseconds to nanoseconds. For example, to take advantage of this feature, Hasbrouck (2019) extends the vector auto-regression and vector error correction models used in the past for econometric analysis of time-series of transaction prices to their high-resolution analysis. The study based on transactions data suggests that in high

resolution, the information share of exchange where the security is listed is substantially higher than its share in trading, possibly due to the presence of designated market makers (distinguishing between the quality of liquidity provision). Second, data of the orders involved in the auctions leading to the transaction prices are available. Example: Brogaard et al. (2019) analyse the contribution to price discovery of market and limit orders by high-frequency traders (HFTs) and non-HFTs. Their study shows that in the case of HFTs, their limit orders contribute more to price discovery than their market orders. When volatility increases, however, price discovery switches from limit orders to market orders. Baruch and Glosten (2019) model limit order markets under imperfect competition, quote cancellations and tail expectations that arise in the markets due to certain rigidities in such markets. Their model includes quote cancellation and tail expectation to deal with a few realities of today's markets.

message_type	order_ref_no	timestamp	Price	Volume	match_number	buy_sell	attribution	REMARK
D	51702192	36007719475135	0	0	0	NA	NA	CANCEL
A	52006412	36007720177996	268.48	100	0	b'B'	NSDQ	BID
A	52012636	36008220480644	268.57	100	0	b'B'	NSDQ	BID
A	52017768	36008444428810	269.31	1	0	b'S'	NSDQ	ASK
A	52027224	36009034636520	268.25	300	0	b'B'	NSDQ	BID
D	51595580	36009035208389	0	0	0	NA	NA	CANCEL
A	52027240	36009035921261	268.38	100	0	b'B'	NSDQ	BID
A	52032800	36009310669992	269.22	30	0	b'S'	NSDQ	SELL
D	52032800	36009310838184	0	0	0	NA	NA	CANCEL
D	51032720	36009310949488	0	0	0	NA	NA	CANCEL
A	52032872	36009310953402	269.35	100	0	b'S'	NSDQ	ASK
D	51994064	36009311087371	0	0	0	NA	NA	CANCEL
A	52032896	36009311808371	269.35	100	0	b'S'	NSDQ	ASK
E	50769360	36009311952448	0	100	1061138	NA	NA	TRADE
A	52032908	36009311977461	269.1	100	0	b'S'	NSDQ	ASK
A	52032916	36009311992873	269.21	50	0	b'S'	NSDQ	ASK
A	52032920	36009311993892	269.26	100	0	b'S'	NSDQ	ASK
E	51077688	36009312015929	0	40	1061140	NA	NA	TRADE
E	51021988	36009312015929	0	10	1061139	NA	NA	TRADE

Figure 1: Snapshot of data from a modern limit order market, Stock: TSLA
Exchange: NASDAQ, Date: 27-03-2019

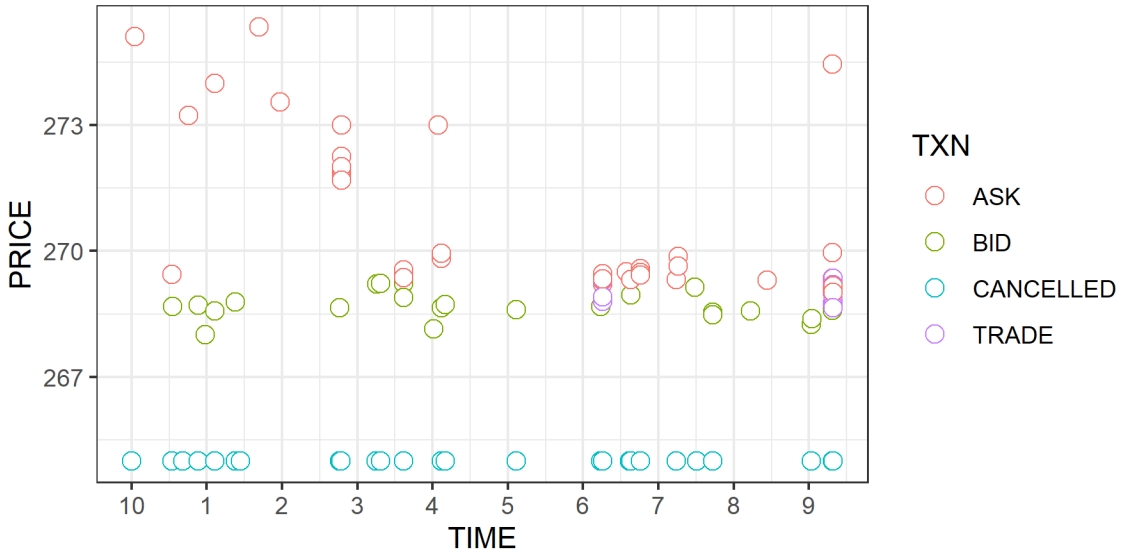


Figure 2 10 seconds of trading activity in TSLA (NASDAQ) starting 10am (large ticks are seconds), 27-03-2019, Price (USD), Time (nanoseconds). Cancelled orders show only the time and not the price

Addressing the realities of today's markets was the call given in O'Hara (2015) to researchers. This dissertation is an attempt to contribute to high-resolution market microstructure research. The collection of essays comprises two theoretical and two empirical contributions. The research outcome applies to the microstructure of any order driven market and helps improve the mechanism and market design. However, the context, illustrations, assumptions, and empirical examples in this dissertation are based on examining intraday trading of the stocks in an equity stock exchange.

O'Hara (2015) summarises the meaning of 'information content' in today's markets as an extension of the traditional understanding of information in the learning models of market microstructure. When the unit of information is 'trade prices,' traders learn from such attributes as 'buy or sell' trades, trade size, number of trades, the time between orders, the time between trades. Extending this when the unit of information is 'orders,' the information from orders is not embedded in the prices alone. An event in the market consists of a tick. Consider ticks to carry an information-set called 'marks' (Price, Quantity, Time of arrival, Buy/Sell/Modify/Cancel), the interaction between which results in the transaction prices. The ticks include the bids, offers, cancellations and trades. Every tick is a step in the auction and involves a price-time sorting of

the order book. Depending on the information we want to study, extend marks to create extended marks. A few such extended marks include cumulative liquidity (with various definitions), order and trade durations, price series for bids/offers and mid-quote price series (after removing staleness or irrelevance to the signal). Apart from the price and quantity, traders execute their information using appropriate quotes (limit, market) and their arrival and departure from the market. Marks contain such information.

Similarly, notions of noise and private information may convey additional meaning today. The source of private information could be from more information about the asset, better analysis of news, market events, non-market events, colocation – split-second advantage with public information, access to information from exchanges or the trader's order flow.

While order driven markets have eliminated designated market makers, researchers continue to invoke their presence in the market and trading models. O'Hara (2015) calls such constructs (the presence of market makers or invisible auctioneers invoked to enforce equilibrium) as 'fictional'. The traders who may, at times, provide market-making services have vastly different characteristics. Besides, the assumption of equilibrium during fast trading markets does not have any theoretical or empirical support. On the other hand, empirical evidence and theoretical support exist to argue why the existence of equilibrium is difficult due to time and communication constraints. This is identified as a key research gap.

To bring market microstructure models closer to reality, in chapter 2 an intraday trading model based on a total order book model is proposed, where trading is done under disequilibrium and in the absence of market maker. The motivation to trade is the need to transform the state of the asset. It uses a systems approach where the trading system attempts to achieve an optimal trading rate to optimise the state transformation. This contrasts with establishing an equilibrium in price and quantities or a game theoretic view of the market, where an equilibrium is set in the trading strategies of the agents. This is a system's view of the behaviour of the traders. The model argues that in high resolution the system's view will dominate over the individual trader. A stochastic optimal control problem is set up and numerically solved. However, requires intuitions on information aggregation in these markets as discussed below and the results from chapter 3 that

deals with the emergence of prices in disequilibrium (another effort to take the model closer to reality).

Chapter 2 also introduces propositions around information aggregation in these markets. A theoretical construct of 'Information content structure' is proposed, which is based on the intuition contained in Easley and Ohara (1992). The result in that paper showed that not only trade transactions but also the lack of trades contains information. In chapter 2 this is inferred in the following way: The market participants are continuously learning from events in the market and this collective information is stored in an Information content structure.

Such a structure mediates information aggregation into the low frequency public signal and not the high frequency transaction prices. Both the transaction prices and the public signal are part of the price system. The public signal is updated with price discovery auctions. Non-strategic traders who do not have any public information trade around the public signal. Strategic traders with private information trade based on the transaction prices.

Existing models also assume the emergence of prices out of an equilibrium or through the services of an invisible auctioneer. Empirical evidence cited in the literature shows that a demand-supply equilibrium in these markets cannot be achieved in the timescales in which the modern order driven markets operate. Hence, it calls for alternative explanation to the emergence of the prices. Chapter 3 in the dissertation proposes and explores the theory of 'auction as a search' as an explanation to the tatonnement in these markets. Using a model inspired by search as used in biology and robotics it is shown that the emergence of prices and the quantity trading, two necessary features of any tatonnement algorithm can be explained by the model. This analysis takes support of diffusion and literature and a reaction-diffusion-advection model. Further, analysis is done using a simulation based on pseudo-experiments to draw deep insights into the mechanism. Thus, search of buyers by sellers and vice-versa is shown to be a potential alternate explanation. Although a deep observation into data emerging from order driven markets with a price-time priority hints at such a plausibility and a search mechanism as intuitive, the results of chapter 3 provide a strong basis to accept the alternate mechanism for tatonnement for high resolution analysis. For analysis done at much lower resolution, time-scales of seconds to minutes the existing demand-supply mechanism is still acceptable. Prices in the intraday trading

model in chapter 2 emerge out of the buyer-seller search mechanism. Thus chapter 2 can explore a high-resolution model by restricting the time of analysis to a single trading day. A conceptual proposition is provided, where the transaction prices emerge around the equilibrium price. The equilibrium price is shown as a global optimum. But time restrictions imply lack of effective communication and hence the system intuitively tries to guess the local optima around the global optima. Hence, the trading rate (which is the control in the chapter 2 stochastic optimal control problem) should move between a high trading rate to a low trading rate. Such movement is facilitated by quote modification of present and incoming orders.

Chapter 4 is an empirical study that explores the role of liquidity in high resolution. It examines if a broad measure of liquidity, such as the liquidity residing in five best bid prices and five best offer prices, has a role in the wealth traded through the trading system. This study is the first high resolution investigation of liquidity in high resolution. It investigates liquidity in a direct way, unlike most studies that use the prices as a proxy. Hitherto, researchers have examined the role of liquidity at the transaction prices and the best bid and best offer prices which is often seen as pseudo-transaction prices. The proposed model considers broad liquidity to have two parts - a liquidity store and liquidity flow (time dimension). Liquidity in the limit order market is modelled as a liquidity store and incoming information from the latent order book, to add or cancel liquidity, is modelled as a liquidity flow. The dimensions covered are depth (five levels in the order book, each on buy and sell side), rate of change in this liquidity on bid and ask side (through order placement and cancellation) and the rate of wealth traded through the transactions.

A fully probabilistic model is employed. The results show, that for the sample of stocks from NASDAQ 100, the liquidity model can explain the wealth traded through the trading system when observed in business time. Estimation is done using vine-copula regression as discussed in Cooke et al. (2019). The explanatory power is satisfactory in high resolution and short periods of business time but weaker in aggregated longer time periods. The interval score is within acceptable measure of good fit in 54.8 percent of the regimes. A regime constitutes of 300 business time units, where a business time unit is defined as the trade of 100 shares. The model fit is low in longer periods of time, which suggests that the precision of the signal is low.

It is observed in individual stocks, the liquidity pool for large part of the trading day goes into a steady state equilibrium. The chapter suggests that when the rate of change of buy side or sell side liquidity is non-zero it coincides with periods of price discovery auctions. In periods of steady state equilibrium, inflow of buy and sell orders through marketable orders and limit orders (including new buyers and sellers) is coordinated with the transaction rates and cancellation behaviour. Algorithmic trading and high frequency trading technology enables precision liquidity management in high resolution. After the price discovery auctions the traders set liquidity conditions that enable wealth transactions at an optimised transaction rate. This provides a partial evidence for the intraday trading model described in chapter 2, where trading was motivated by a need to transform the state of the asset held with the trader.

The liquidity pool is not in steady state during early trading hours and periods of price volatility. A probable inference suggested is that these are periods of price discovery auctions and are evidence of higher quality of provisioning by market makers (as discussed in Hasbrouck (2019)). This interpretation contrasts with Brogaard et al. (2019), who study the case of HFTs and, suggest that when volatility increases the price discovery moves from limit orders to market orders. The observation of the liquidity variables across our sample stocks suggests a role of the broader liquidity during the price discovery auctions.

This chapter also contributes in methodology as used in market microstructure research. The proposed model is a fully probabilisitic model (bayesian hierarchial model). The model is trained and used to forecast using Vine Copula Regression as discussed in Cooke et al. (2019). In this way it attempts to expand the microstructure toolbox.

The empirical study in chapter 4 required a new paradigm for data handling which has been provided in chapter 5, a joint work with Ranjan Chakravarty. Firstly, the chapter reports discovery of stochastic processes in data from order driven markets that is different from microstructure noise. This noise, named as non-microstructure noise, needs to be clean for effective estimation in high resolution. Secondly, the chapter contributes to the challenge of providing a rationale to work with quotes (orders) and link them to prices. It is a challenge because many orders eventually get cancelled without getting traded. Kichia and Protter (2015) provide the theoretical framework to consider the orders within our existing understanding of

evolution of prices. They do so by using the technique of expansion of filtration. Chapter 5 operationalises this by providing a method to check if any order (quote) actually participates in an auction (Brownlees and Gallo (2006)). The result is a data paradigm to handle data for high resolution analysis. This paradigm extends Brownlees and Gallo (2006) and Brandorff Nielsen et. al (2009) to identify non-microstructure noise using limit order book modelling and measures of central tendencies. It provides a flexible approach to researchers working in high resolution studies to adapt to their unique requirements. The efficacy of the data paradigm is illustrated with an empirical example.

The contribution of this dissertation can be summarised as follows:

1. It *extends* the research area of high-resolution market microstructure, adding theoretical models, empirical results, and methodological contributions.
2. *Theoretical Modelling*: The Intraday trading model incorporates analysis at high resolution. The model is closer to real order driven markets in two aspects. First, analysis in disequilibrium, aided by an alternate mechanism for the tatonnement in these markets. The need for prices to emerge through an equilibrium of supply and demand or impractical assumption of the presence of an invisible auctioneer is relaxed. The objectives of the tatonnement mechanism are achieved through 'Auction as a Search.' Second, the absence of market maker, like 'Smooth Trading' model of Kyle et al (2018), Vayanos (1993), Du and Zhu (2017), where this dissertation additionally extends the analysis to high resolution. Thus, it differs in the implicit time dimension and other aspects necessitated due to this change. The time in the latter model is to achieve equilibrium. The time in the intraday trading model is restricted to a single trading day due to high resolution analysis. There are also propositions to explain information aggregation in order driven markets.
3. *Alternative explanation to tatonnement mechanism*: Auction as a search of buyers by sellers and vice-versa is established as a probable tatonnement mechanism. It provides the opportunity to researchers to consider building more realistic models for high resolution studies.

4. *Information content of quotes*: A unique high resolution empirical study into the role of broad liquidity is presented. Broad liquidity provides an explanatory role into the wealth traded in the markets. This demonstrates the information content embedded in quotes and a technique to examine it. The results show how traders equipped with fast trading technology execute their mandates by utilising the limit order book and the latent order book. There are indications that state transformation of the asset held, may be one of the goals in trading and that trading system view of the market may be one way to look at markets in high resolution and in business time.
5. *Contribution to enhance the microstructure toolbox*: Data handling paradigm for high resolution orders data, application of probabilistic Bayesian hierarchical models, application of pair copula construction and vine copula regression to high resolution data and market microstructure research.
6. *Experimental techniques*: The pseudo-experiment simulation in chapter 4 demonstrate a technique to simulate zero intelligence models of order driven markets. It enhances our understanding of the relative performance of brownian, superdiffusive and ballistic search. This is expected to be useful in mechanism and market design especially for assets with lesser trading interest such as small and mid-cap stocks.

CHAPTER 2

INTRADAY TRADING IN ORDER DRIVEN MARKETS

This chapter investigates, in high resolution, the information content present in the orders and trades and information aggregation in an order driven market. A continuous time intraday trading model is described for a market that trades a single risky asset and employs an order driven market mechanism with the usual price-time priority. The key features of the model are listed as under:

1. Trading is motivated by a need to change the state of the asset held (risky asset to risk free monetary holding and vice-versa).
2. The model is characterised by a horizon of a single trading day, absence of market makers and trading in disequilibrium.

Price System

3. Price emerges as a result of search of buyers by sellers (or vice versa) (tatonnement mechanism described in chapter 3). The market clearing condition is relaxed. This enables an intraday trading model in high resolution. The model works with locally optimal solutions and does not look for the global equilibrium.
4. Using a series of propositions, a method of information aggregation in these markets is proposed. The price system is considered to contain the set $\{Y, y_p\}$; Y is a random variable - the transaction prices, and y_p is the public signal of the asset price. The public signal is the price considered as the asset price by many active market participants and is conditioned on the available public information. The public signal is not updated intraday in this model.

Information Aggregation

5. Information aggregates into the public signal y_p at a low frequency (and not the price Y that evolves at a high frequency). This aggregation is facilitated by a conjectured Information Content Structure, a mechanism that holds the collective market memory. Information content structure (ICS) is defined as the total price shock communicated by the agent in unit time. A quantification of the ICS has been provided.

Agents / Traders

6. There are two kinds of traders in the market; Traders with private information (Strategic traders) use the emerging asset price as the signal and traders without private information (Non-strategic traders) trade on the public signal. All traders can choose to deploy a limit or market order to execute their trades. There exists a minimum number of non-strategic traders who deploy market orders to avoid the no-trade equilibrium due to non-

availability of market orders. To conceive this a subset of the non-strategic traders is considered to continuously trade in the market by alternatively placing a limit order and market order. This subset is referred to as the non-strategic noise traders. Unlike classical noise traders, the non-strategic noise traders try to make a non-negative profit during a trading day.

7. The presence of strategic traders depends on the price and time. Strategic traders come into the market only when price is right. The price is the one-sided boundary for strategic traders. Only a part of the Total or Latent order book reveals as limit orders in the visible order book.

Losses

8. Who absorbs the losses in the absence of market makers is critical. Strategic traders bear losses and other transaction costs in the trading system. The intuition is that they would like to actively manage the risk of non-execution or the time value of money. Another implication is that price resets, to increase the trading rate, can be expected from strategic traders, or, they may try to cut the impact cost during trades. Nonstrategic noise traders are expected to be wealth neutral in this model. They provide the bridge to the non-synchronous arrival of buyers and sellers in the model in the absence of market marker.

The System view and Utility functions

9. The trading system should offer the traders an optimal outcome. The system can optimise the outcome by setting the appropriate rate of transaction. The utility function modelled is the utility of the system. Individual agents have their own utility functions but it is not modelled. The trading system is influenced by the system level dynamics and utility, which is simply the aggregate of the individual behaviour. Bone (1998) showed that if we assume the individual utilities of traders are CARA, then the collective choices that emerges is EU.
10. A source of error emerges from the fact that traders have access only to the visible order book and not the latent order book that has to be inferred. Trader quotes arrive at different points in time and the system would need the transaction rate as a control. If the

transaction rate does not meet the expected trading objective of the traders it leads to quote modifications.

11. Intraday trading can be studied based on aggregate behaviour of the traders. When a snapshot of the continuous and dynamic trading process is taken, encompassing a day of trading, the high frequency and price priority nature of the market design ensures that the trading system characteristics dominate over other important criteria such as dividend signals, consumption beliefs etc. Such an argument may not hold in longer periods.
12. The wealth in the limit order book is analysed. It is assumed that dividends are neither paid nor announced during the analysis.

Wealth traded used with transaction rates- A market observable that includes time

13. Market clearing and transaction rate or rate of trades are concepts of business time. In the context of continuous auctions, market clearing is a weaker notion as it does not involve time. Since, market clearing involves both quantity and price, it can be conceptualised in terms of wealth used with transaction rates which is a market observable.

Dis-equilibrium or lack of competitive equilibrium

14. Figure 3 shows that while equilibrium trading is expected, trading is more likely to follow non-equilibrium path due to the lack of time for the information flow required for the equilibrium. In high frequency markets, market clearing will not be met transaction to transaction. It can only be established in coarser time periods. To build predictability into trading, traders would want to see the rate of wealth traded to be constant. If the rate of wealth traded is not constant, they need to adapt their trading strategy to the prevalent scenario. Thus, $\frac{d}{dt}QP = constant$. However, such a wealth transaction rate is specific to the trader. Hence, there will be heterogenous beliefs in $\frac{d}{dt}QP$. Traders would want accumulation or liquidation of stocks leading to constant change in wealth in their own books. **An observer of the trading system would however see that the system will try to go towards optimisation of the wealth transacted by controlling the trading rate. This is the equivalent of an equilibrium in the trading system view of the market.**

Optimal rate of transactions

15. What is the optimal rate of transactions at which the market clears? Since transactions take place in disequilibrium the objective of the trading system will be to minimise the error from the equilibrium so that it is as close to the market clearing line as possible. The system will attempt to reduce the variance of the wealth transactions in actual versus equilibrium prices. Since we do not know the equilibrium trading rate, the system will vary from slow trading rate to faster trading rate. That is, from low quantity traded to large quantity traded.
16. Equilibrium will be transitory in this model because of the assumptions. The assumption in the intraday model is that when the time horizon increases, other signals such as dividends and investor's utility come into play.
17. The system controls the trading intensity to control the wealth traded through the system. The proposed representation recognises the ability of participants in such markets to observe market events and calibrate their quoting activity.

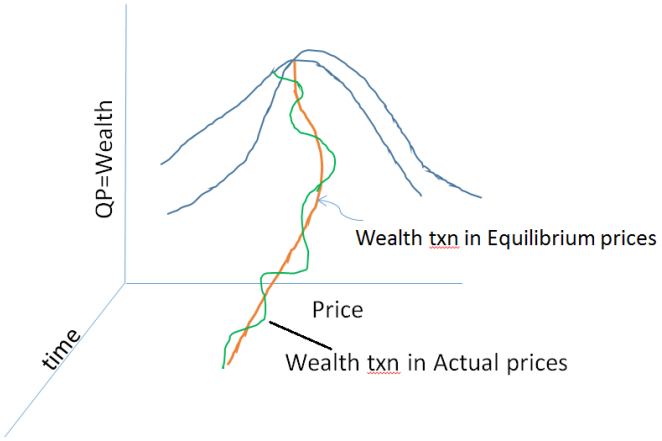


Figure 3: When trading in disequilibrium, the path of trading fluctuates around the equilibrium path (Illustration)

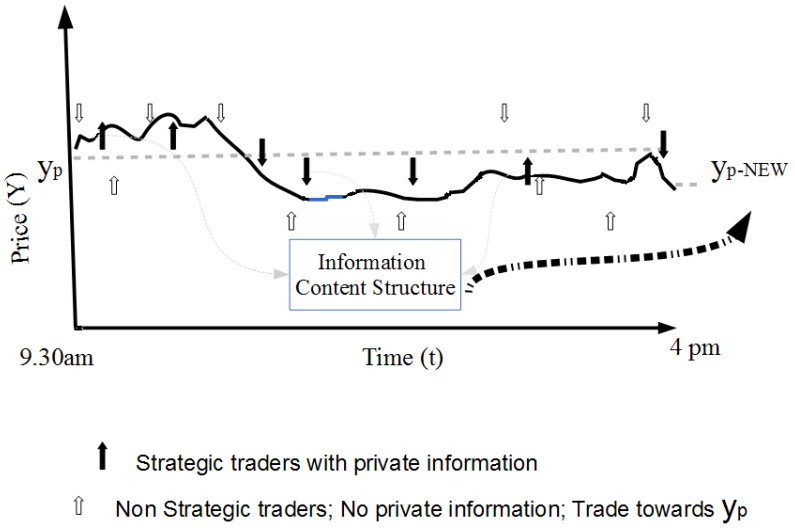


Figure 4: The Set up illustrating the presence of strategic traders and non-strategic traders. The public signal at the start of the day is yp that gets revised at the end of the day. Strategic traders quote with reference to the price and non-strategic traders refer the last traded price and move in the direction of the public signal yp . The market records the volume-weighted price shock of all traders in the ICS.

THE MODEL

The Intraday trading model is set up as a stochastic optimal control problem where the trading system attempts to optimise the rate of trading to control the wealth (monetary value of asset changing its state) traded through the system.

The utility function of such a system is assumed as (1)

$$U\left(\pi_c(s)\right) = -e^{-A\pi_c(s)} \tag{1}$$

Here π_c is the wealth that is converted or transformed or asset that has changed state. In a time interval dt the wealth that is converted is the transaction rate ¿Y , where $Y(t)$ is the price of the asset bought or sold.

Wealth in the LOB

Let $X(t)$ denote the wealth in the visible limit order book, $S(t)$ denote the inventory of the asset in the LOB and $M(t)$ denote the monetary cash holdings. The wealth at the beginning and end of the day is $X^0$and X^1respectively.

During the day, the wealth in the LOB is $X(t)=S(t)*Y(t)+M(t)$. There is addition and deletion of inventory and monetary holdings through the process of quoting, transaction (state conversion & subsequent withdrawal) and cancellation of quotes. (refer (2))

$$dX(t)=X_t\left[-rdt+\phi dt-\phi C_D dt-\delta dt+w_t dt-C_D w_t dt\right]-\pi_c dt-\pi_c C_T dt+\sigma_w dW \qquad (2)$$

Where the notation denotes the following:

ϕ : The fraction of the wealth that is reinvested or redeployed immediately after conversion. This comes from the Non-strategic noise traders.

π_c: Wealth converted or wealth whose state was already transformed. This is the control in the model.

C_D, C_T :C_D is the transaction cost incurred per unit of wealth during deployment and C_T is the transaction cost incurred during the transaction.

δ: δ is quote cancellation. Some of these may return through quote modification but is captured in the parameter w_t.

w_t : Wealth getting added through quotes coming from latent to visible LOB.

To simplify (2) consider the total cost on the system other than state conversion as: Cost= Transaction cost + Time value of money + risk of non-execution

$$C(t)=\left(r+\phi C_D+w_t C_D\right) \qquad (3)$$

Time value of money and risk of non-execution influence the quoting behaviour - the behaviour of cancellation and modification at trader level. Cost is one of the drivers of any trading process. From the perspective of the trader, the cost is reflected in the tatonnement process and hence indirectly into the trading system. The perspective of the system is depicted below:

The assumption is that while individual traders react to the slower or faster transaction rate that affects their time value of money or risk of non-execution, it can have a system level effect only after a certain threshold is crossed. The net addition of wealth factor in terms of inventory being

offered or bid as $I(t)$, combining parameters ϕ, w_t . The conversion is stated as s , where λ is the trading intensity and s is the averaging trade size that is assumed constant for the system. Thus, the system controls through the change in the trading intensity. (3) is obtained by simplifying (1) using (2).

$$dX(t) = -CX_t dt + I_t Y - \lambda * s * Y(1 + C_T) + \sigma_w dW \tag{4}$$

Let X^C denote the wealth whose state has been converted and X^R denote the wealth that awaits conversion. This wealth will have exposure to the risk emerging from the public signal if the conversion does not take place during the trading hours. In (3) note that the valuation of π_c is based on the transaction price Y. X^R evolves during the day and once the new public signal is ascertained, it transforms as X^1 at the end of the period after incorporating the ICS and the model assumption (restriction) to keep the non-strategic noise traders wealth neutral. Hence, the discrete process $y_p^0 \triangleright y_p^1$ incorporates market memory (ICS) as well as adjusts to incorporate the trading costs and losses of non-strategic noise traders.

$$X^c = \int_t^T \pi_c(s)\, ds \tag{5}$$

$$X^R = X^0 + \int_t^T dX \tag{6}$$

The wealth awaiting conversion can be better modelled as

$$dX^R = dX - \lambda_t * dY^{\square} \tag{7}$$

The problem is to maximise the utility of the trading system.

$$V = J(\pi) = \max_{\pi(s)} E\left\{ \int_{s=t}^{\infty} -e^{-A\pi_c(s)} \right\} = \max_{\pi(s)} E\left\{ \int_{s=t}^{T} -e^{-A\pi_c(s)} + r(X^1) \right\} \tag{8}$$

Equation (8) is interpreted in the following way: Since the purpose of the trading system is state conversion, its utility for a single day of trading is given by the first term. And the second term gives the utility for the future trading days. Non optimal trading on any day shall impact X^1 by the change in the public signal of price or the inflow and cancellations in the subsequent periods. (8) is subject to the following constraints.

Wealth constraint defined by the process $dX(t)$ given in (4). The asset price Y defined by the process (9) and public signal defined by the process (10).

$$dY = \alpha_1 Y dt + \sigma_Y d B_1 \tag{9}$$

$$d y_p = \alpha_2 y_p dt + \sigma_p d B_2 \tag{10}$$

(10) is assumed to be an approximation of the signal observed due to the discrete process that gets generated basis the market memory (ICS). i.e $y_p^0 \triangleright y_p^1$ conditioned on the ICS. In the current model ICS becomes zero after the preauction adjustment while the public signal in (10) is a continuous time adjustment.

Value function (equation 11) of the form $V(X, y_p, Y)$ is conjectured and verified.

$$V(X, y_p, Y) = -\exp\left(\psi_0 + \psi_w X + \psi_p y_p + \psi_Y Y\right) \tag{11}$$

The corresponding HBJ equation becomes (12):

$$0 = \max_{\pi(s)} \left\{ \frac{-e^{-A\pi_c(s)} + r(X^1)}{V} + \psi_w \left(-C X_t + I(t) Y - \lambda s Y (1 + C_T)\right) \right\} + \psi_Y \alpha_1 Y + \psi_p \alpha_2 y_p - \frac{1}{2} (\tag{12}$$

The optimal control for above HBJ equations is obtained as (13):

$$\pi_c^{\square} = \frac{-1}{A} \log \frac{\psi_w (1 + C_T) V}{A} \tag{13}$$

Putting back $\pi_c^{\square}$ into (12) yields (14):

$$0 = \frac{-\psi_w (1 + C_T)}{A} + \frac{r(X^1)}{V} + \psi_w \left(-C X_t + I_t Y - \lambda s Y (1 + C_T)\right) + \psi_Y \alpha_1 Y + \psi_p \alpha_2 y_p - \frac{1}{2} \sigma_w^2 \psi_ \tag{14}$$

The coefficients of variables are separated and equated to 0. A system of equations is obtained that do not have any closed form solutions but can be solved numerically for the unknowns $\psi_0, \psi_w, \psi_p, \psi_Y$.

Figure 5 plots the wealth in such a trading system during a typical day. Additionally, for the simulation a specific relationship between the trading rate and quoting rate given by n has been assumed. The system's wealth ideally should end in negative showing the cost of

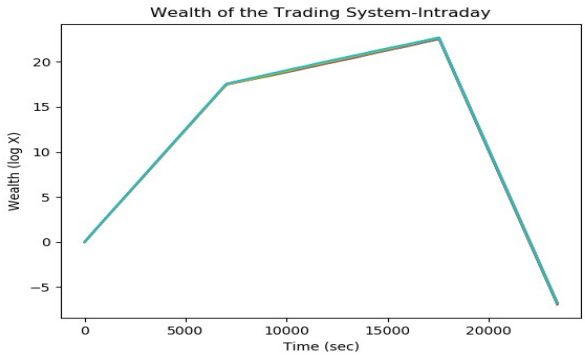


Figure 5: Wealth ($logX$) of the trading system as it evolves through the trading day 9.30 am to 4.00 pm as per (8). The values of the constants C = 7.69231e-7, C_t= 1.19e-9, σ_W = 6.5372e-4,σ_Y = 3.2686e-4 , σ_P = 6.5372e-5, α_1 = 4.27e-7, α_2 = 4.274e-5, λ = 4;s = 100, n = 1.0025, n = 1.0005 and n = 0.995 in the first 30 percent of the trading period, 30-75 percent and the remaining trading period respectively. The plot shows 10 simulations and it remains identical even for a large number (>1000) of simulations.

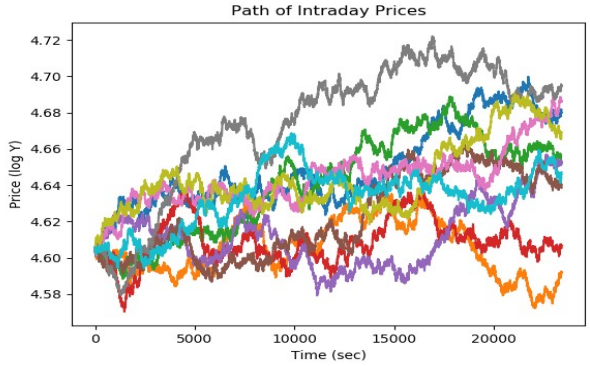


Figure 6: Sample path of intraday log-prices as per (13). The values of the exogenous constants C = 7.69231e-7, C_t= 1.19e-9, σ_W = 6.5372e-4,σ_Y = 3.2686e-4 , σ_P = 6.5372e-5, α_1 = 4.27e-7, α_2 = 4.274e-5, λ = 4;s = 100, n = 1.0025, n = 1.0005 and n = 0.995 in the first 30 percent of the trading period, 30-75 percent and the remaining trading period respectively. The plot shows 10 simulations and it remains identical even for a large number (>1000) of simulations. The initial value of the asset price assumed is log(100).

transaction which in this model is borne by long term strategic investors. Figure 6 gives the sample path of intraday prices. This is governed by the SDE (14). Figure 7 gives the trading intensity with different precision of the signals. In optimal trading the trading intensity shall

remain stable if we assume an average trade size (for example 100 shares per trade). In our model the precision of the longer-term signal is a constant in intraday and changes before the next trading day. When the precision of the signal from the intraday prices is low (i.e the volatility or variance is high) it results in lower trading intensity. Only when the precision of the intraday price signal is high, the trading intensity is higher. Thus, one way the trading system can increase the trading intensity is by increasing the precision of the intraday trading prices. In a period where the trading intensity is not satisfactory to the traders, it may call for modification of quotes to reduce the variance (increase the precision of the signal).

In Figure 8, the relationship between the trading activity and the quoting activity in terms of the inventory I) is described. Intuitively, the highest rate of trading can be achieved in a hypothetical system where incoming market orders liquidate the exisiting limit orders. This is depicted in the lower values of n. However, every trading system will settle down to its characteristic trading intensity with a diminishing influence of the quoting multiple.

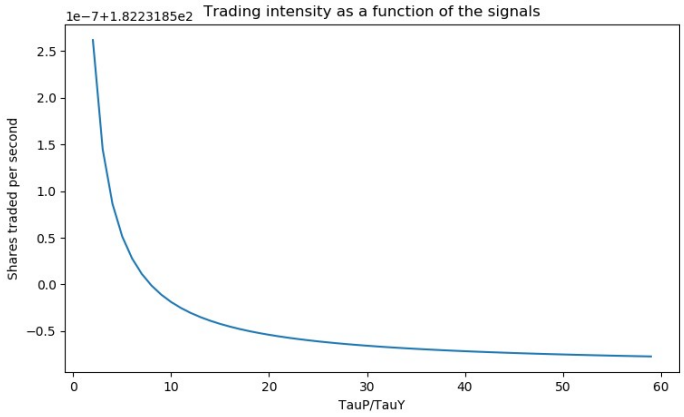


Figure 7: Shares traded in optimal trading with the precision of the signals. The long term signal remains constant through out the day in our model. The values of the exogenous constants C = 7.69231e-7, C_t= 1.19e-9, σ_W = 6.5372e-4,σ_Y = 3.2686e-4 , σ_P = 6.5372e-5, α_1 = 4.27e-7, α_2 = 4.274e-5, λ = 4;s = 100, n = 1.0025, n = 1.0005 and n = 0.995 in the first 30 percent of the trading period, 30-75 percent and the remaining trading period respectively. In the x-axis, $\tau_P = \left(\frac{1}{\sigma_P^2}\right)$ the precision of the public signal which is high and does not change during the trading day and $\left(\frac{1}{\quad}\right)$

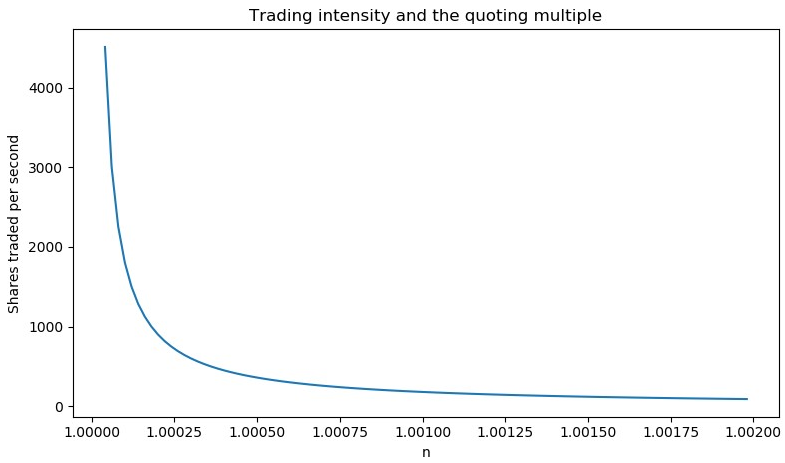


Figure 8: n describes the relationship between the net addition of wealth in terms of inventory (quoting activity $\left(I_t = n * \lambda * s\right)_{\square}$) and the traded assets (y-axis)$\lambda * s$. n is the multiple. The values of the exogenous constants C = 7.69231e-7, C_t= 1.19e-9, σ_W = 6.5372e-4,σ_Y = 3.2686e-4 , σ_P = 6.5372e-5, α_1 = 4.27e-7, α_2 = 4.274e-5, λ = 4;s = 100, n = 1.0025, n = 1.0005 and n = 0.995 in the first 30 percent of the trading period, 30-75 percent and the remaining trading period respectively trading prices relatively lower.

CHAPTER 3

A THEORY OF 'AUCTION AS A SEARCH' IN SPECULATIVE MARKETS

Under conditions where a designated market maker or an assumed auctioneer is not present, this chapter explores 'Auction as a search' as an alternative to the supply and demand paradigm for tatonnement in order driven markets. The tatonnement process is modelled as a search by buyers for sellers and vice-versa.

The walrasian tatonnement is in effect an algorithm wherein an invisible auctioneer has all information regarding supply and demand schedules and comes up with the price that optimises the quantity traded. This emphasises an important function of the market, that of optimising the traded quantity. Any alternative algorithm to be useful needs to meet these two goals. 'Auction as a search' meets both the above stated objectives.

Intuitively, when the principle for matching trades is based on 'price-time priority' in a continuous double auction mechanism, it resembles a search. The need for the model arises as existing theory based on supply and demand assumes the existence or emergence of a walrasian

equilibrium. The above assumption takes the underlying model away from reality when considering intraday markets that trade assets in high frequency through continuous auctions. In high resolution the assumption of an equilibrium of demand and supply leading to optimised quantity traded and emergence of price is fictional. This is especially true in case of order driven markets without designated market makers. The general belief that in financial markets given that trading is continuous, prices can quickly adjust to clear the market is valid but the process is not instantaneous. Information flow is not seamless and frictions exist. Traders are not always present in the market and alternate opportunities exist for the traders. Market makers and dealers in quote driven markets were the key to push the trading to equilibrium.

A total order book model (Figure 9) comprising limit orders and latent orders, in the absence of a market maker is proposed. A zero-intelligence approach of agents is employed using a diffusion drift-reaction model, to explain the trading through continuous auctions (price and volume). The model is inspired by search in biology and robotics. The search (levy or brownian) for model inspired from foraging, chemotaxis, and robotic search. This is a unique way of looking at

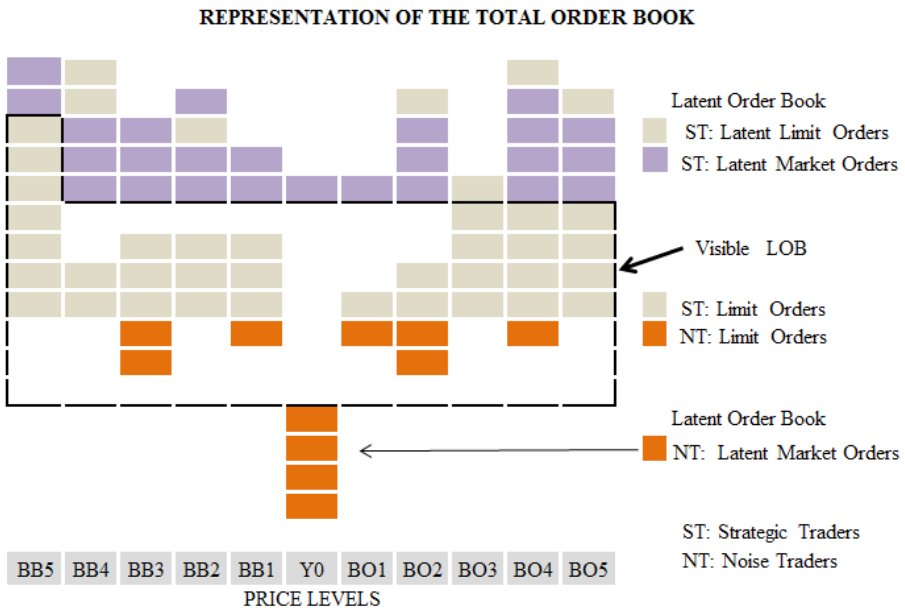


Figure 9: Representation of the Total Order Book as a one dimensional lattice. It comprises the visible limit order book and the latent order book. The dark border represents the visible limit order book given in fig 1. The limit orders from strategic traders (ST) are in grey. The limit orders from noise traders (NT) are placed immediately after a market order is executed by a NT. Strategic traders can place limit orders (grey) or market orders (purple) in future and hold their orders in the latent order book. Noise traders place market orders at the last transaction price Y0. BO-Best Offer; BB- Best Bid.

predator-prey relationships (as either the buyer or seller could take the role of the predator or prey). Only when there is a bias in the system and an imbalance favouring either the buyer or seller, there will emerge a true predator-prey situation.

The case without volume bias: Diffusive case

Exisiting literature in diffusive (levy) search (Zaburdaev et al. (2008)) is adapted by considering both time and velocity as random variables. The quoting behaviour of agents is modelled by considering that the return expectation (distance travelled by an order in a grid) is equal to time spent by the order in the market (flight time) and the (assumed) velocity provided to the order by the agent. Flight time and velocity are the two basic and independent random variables of the model. Dimensional reduction is achieved by modelling only market orders as market orders need to be involved in every transaction. With the above assumptions and deploying existing mathematical apparatus available in diffusive (levy) search literature the tatonnement (evolution of prices and transaction) is modelled. The rate of trades is used to measure the efficiency of the search process. However, lack of bias (due to volume imbalance between buy side and sell side) is a strong assumption. This is later relaxed.

The Complete Model - Search for trades in presence of bias

The stochastic evolution of prices (diffusion+jumps) is modelled using diffusion-advection-reaction model with behavioural dimensions built into it.

Let $g(x,t)$ represent the density of market particles at a price x and time $(0 \le t \le \infty)$. The fundamental set up is given by the partial differential equation (15):

$$g\frac{\partial g(x,t)}{\partial t} = ¿$$

$$\frac{\partial}{\partial x}\left(D(t)\frac{\partial g(x,t)}{\partial x} - \lambda_1(t)\frac{\partial f(x,t)}{\partial x}g(x,t) - \lambda_2(t)g(x,t)\right) + v(t)g(x,t) + q(x,t) \qquad (15)$$

The terms in the right hand side of the above equation are explained below:

- Diffusion (first term): This is the primary search mechanism through which price continuity and quantity transport takes place. In the absence of bias this will remain the only mechanism.
- Drift due to target density (second term): A formulation widely used in foraging studies is implemented (Grunbaum (1998)). In the set up of those studies, the drift velocity is proportional to the gradient resource density. And the coefficient is the taxis coefficient. The use of resource density implies the dimensionality of the model could increase. Hence, in this model the drift resulting from the response to increase in density of targets or consequence in market particles is modelled. $f(x,t)$ is the perceived resource density basis the actions in the system. In this innovation, the conjecture is that, the drift resulting from a gradient of resource density should have a mirror image in the searcher. This mirror image may be different than the original gradient, but is more relevant as it is likely to be a tactical response in the absence of complete information (since the visible order book is not the complete order book). We may err in magnitude but not direction and the system will try to iteratively guess the true gradient of resource density. A time-evolving coefficient λ_1 is used.
- Drift due to movement within the market order particles (third term): It is envisaged that there could be several scenarios where a few market particles could get activated into directed activity resulting in a drift. For example, high transaction activity on one side of the market, could result in noise traders getting released from limit orders suddenly due to the reactions. This early release increases returns for noise traders. If these traders sense the action to continue, they would rush to the other end to repeat the cycle. Similarly, traders yet to place orders may anticipate change in resource gradient and resource density due to news flow and react by adjusting their own positions. λ_2 is a time varying coefficient .
- Net addition to market particles density v (fourth term): $v(t)$ is a time varying composite coefficient of the new order additions v_a, cancellations v_c and transactions (reactions) v_r. $v(t)g(x,t)=v_a(t)g(x,t)-v_c g(x,t)-v_r g(x,t)$. New order additions of market particles includes market particles, hitherto latent, coming into the order book, new traders, limit orders that are cancelled and placed as market orders as the traders become impatient. Hence, this can be a non-trivial component. Cancellations and transactions remove

market particles of strategic traders from the order book. Noise traders move out of market only at end of their flight time.

- q (fifth term) is a source of market particles much larger than what comes in during the average trading day for the asset. It could be a demand or a supply of liquidity. The source may be a single or few point sources or distributed across a price range and even in time.

In summary, a comprehensive model of stochastic evolution of market particles is built. This equation is solved analytically using the Greens function method. The solution obtained is (16):

$$g(x,t)=\beta_1(t)\int_0^t\int_{-\infty}^{\infty}\frac{Q(\chi,\tau)}{\sqrt{4\pi D(T-\zeta)}}\exp\left(\frac{\left(\frac{x}{\beta}-\int_0^t\left(\frac{\lambda_2}{\beta}+\frac{\lambda_1\phi_2}{\beta}\right)ds-\chi\right)^2}{4D(T-\zeta)}\right)d\chi d\tau+\beta_1(t)\int_{-\infty}^{\infty}\frac{}{\sqrt{4}} \quad (16)$$

Here, $\chi=\int_0^{\tau}\frac{1}{\beta^2(s)}ds$ and the initial condition $(x,0)=G_i\omega(x)$.

The significance of β.

The dimensionless quantity β has a physical meaning. On transforming the domain from (x,t) to (X,t_0), each point on the one dimensional lattice is now a space-time variable. The advantage of this construction is that X alone can explain the time evolution of the system. This intuition is used in the following way: β relates the diffusion in the new domain to the original domain.

Since, β is defined in terms of the relative contribution from the net addition to the density of market particles in the two domains, the inference is that at any particular price in the domain (X,t_0) we are able to witness the impact of the deposition, cancellation and transaction over time. Thus, the behaviour is that in one domain v_1 already incorporates time evolution, while in the other v must be integrated over time.

Analytic and asymptotic analysis is provided for several scenarios and examples. The list includes: a steady state solution (where in the square root law is recovered), a large demand, high

reaction rate, drift with no time dependence and low reaction rate, continuous spatial inflow of market orders that is not time dependent and at the limit of continuous auctions.

Figure 10 confirms that in an ideal best search scenario the density exists. As the spread increases, the density is reduced. The interauction times in which the density of market particles exists is analysed and a relationship between the diffusion coefficient, interauction time, initial density of particles and spread is found to exist. For the spread less than 0.2 USD and inter-auction time, 0.001 seconds or lower, less than 40 percent of the existing market particles can trade. There exists a relationship between the spread and the diffusion coefficient for an efficient market. So, markets that also depend on drift (resulting from high transaction rates involving noise traders) apart from diffusion are adversely impacted when the spread is high.

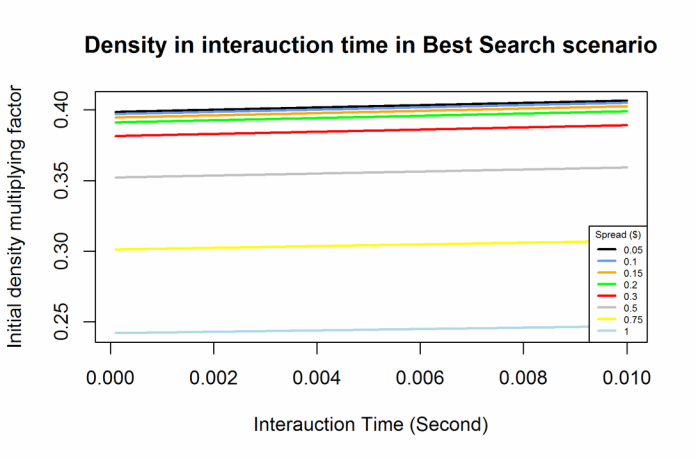


Figure 10: The multiplying factor to the initial density as a function of the interauction time and spread (B).

Numerical simulation

Numerical simulation of the model extends the understanding of the relative performance between brownian, superdiffusive and ballistic search in the model.

A numerical simulation of the model was performed to improve the understanding of the results. This numerical simulation is different than the regular reaction-diffusion simulation because while orders may be present in the total order book, they come into the limit order book at different times. The order arrival into the limit order book is modelled using a poisson process. The numerical simulation run on the model is based on a Pseudo-experimental set up. It brings

out the search efficiency of different search regimes (ballistic, superdiffusive and brownian) in the simulated set up. The rate of trades that emerges from trade duration is a natural candidate to measure the efficiency of search. When biases exist due to resource density, the superdiffusive search performs better than brownian search. In the absence of bias brownian search is equally efficient in high resource density. In low resource density ballistic search performs the best followed by superdiffusive search. (refer figure 11 to figure 13). Performance of ballistic search is not affected by resource density or the presence of bias. (figure 14). Additionally, figure 15 gives the behaviour of prices under the model.

An acceptance of this proposed alternative should help researchers, working on price discovery and trading models in high resolution, build models that are closer to reality.

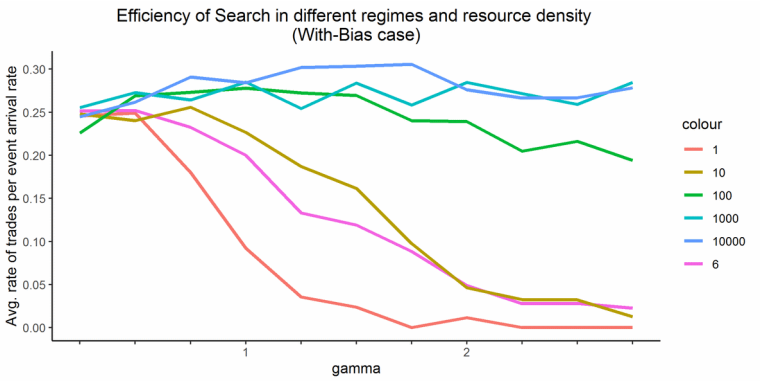


Figure 11: The efficiency of search depends on the search regime and the resource density. The former is set through the quoting and cancellation behaviour. And the latter through the event arrival rates. Additionally, the presence of Bias due to resource density increases efficiency of superdiffusive search more than brownian search. In low resource density, superdiffusive and ballistic search perform better.

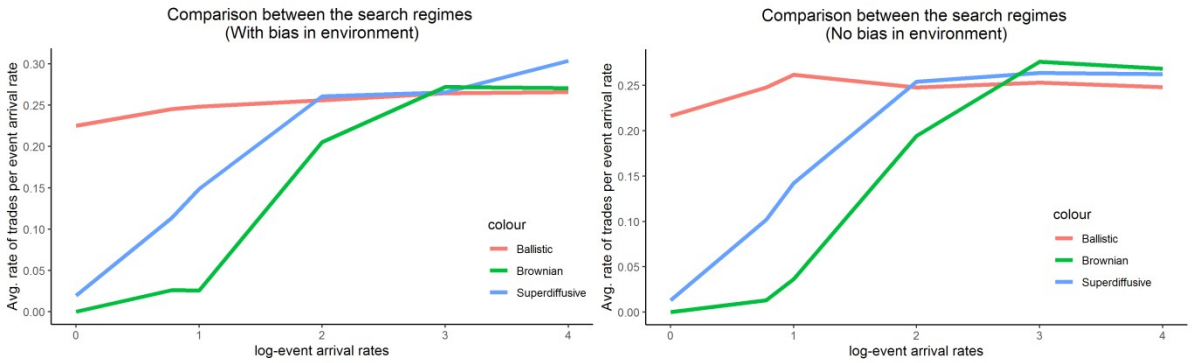


Figure 12: Comparison of the search regimes in different resource density and influence of bias. Inthe presence of bias, superdiffusive search is more efficient than brownian. Ballisticsearch performs the best in low resource density and is less influenced by resource density

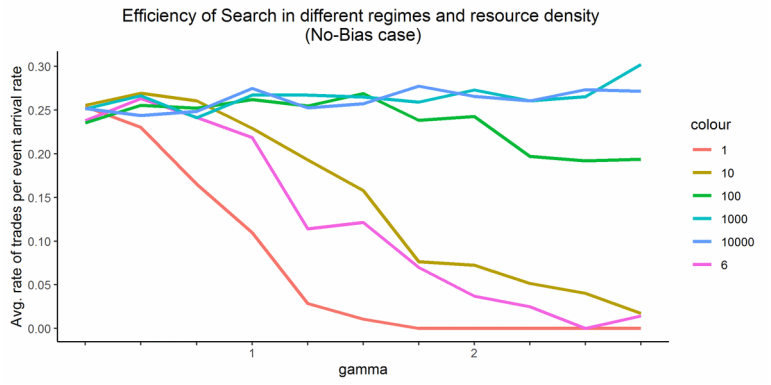


Figure 13: The efficiency of search depends on the search regime and the resource density. The former is set through the quoting and cancellation behaviour. And the latter through the event arrival rates. In low resource density, superdiffusive (1< <2) and ballistic search(<1) perform better. When resource density increases and in absence of any biases in the environment, brownian search is as efficient as the others.

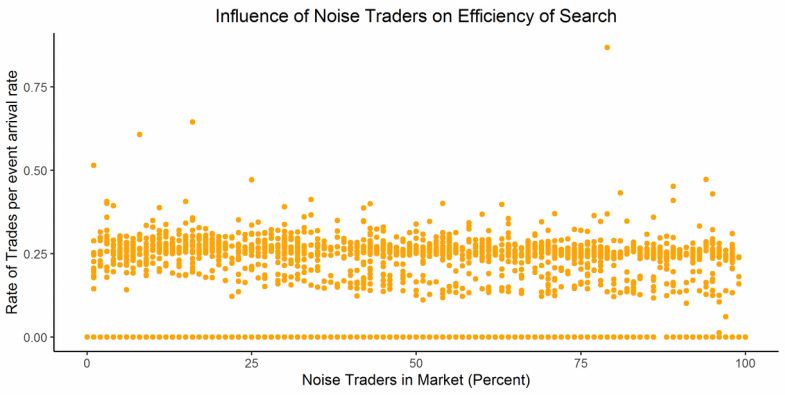
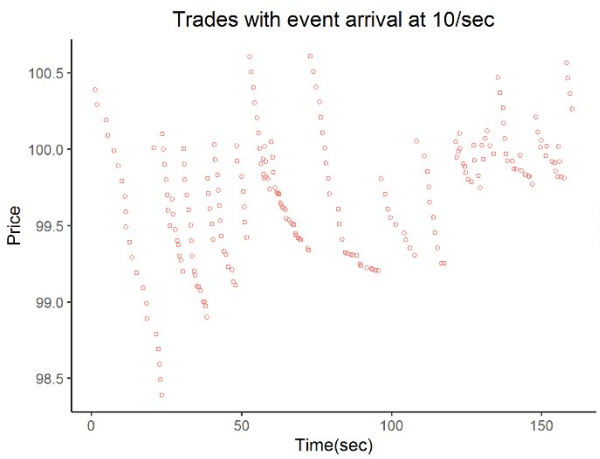
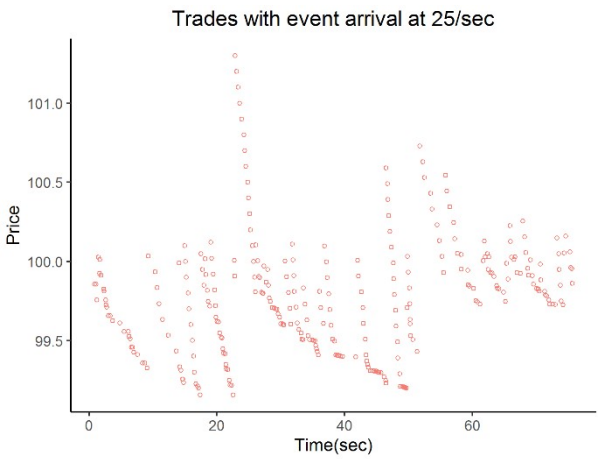
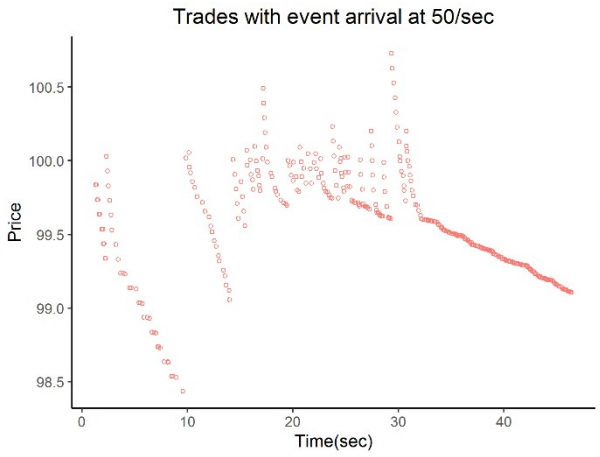
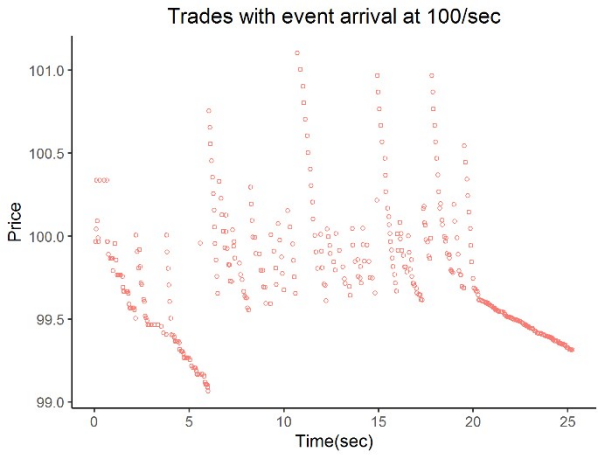


Figure 14: The percentage of noise traders in each trial of the simulation is randomly assigned. The number of noise traders does not affect the efficiency of search. The possible reason could be the poisson arrival of the traders. The above is the no-bias case that plots the rate of trades per event arrival rate with the percentage of noise traders in the market. The plot of the with-bias case is also similar.


Trades with event arrival at 10/sec
Price
Time(sec)
100.5
100.0
99.5
99.0
98.5
0
50
100
150


Trades with event arrival at 25/sec
Price
Time(sec)
101.0
100.5
100.0
99.5
0
20
40
60


Trades with event arrival at 50/sec
Price
Time(sec)
100.5
100.0
99.5
99.0
98.5
0
10
20
30
40


Trades with event arrival at 100/sec
Price
Time(sec)
101.0
100.5
100.0
99.5
99.0
0
5
10
15
20
25

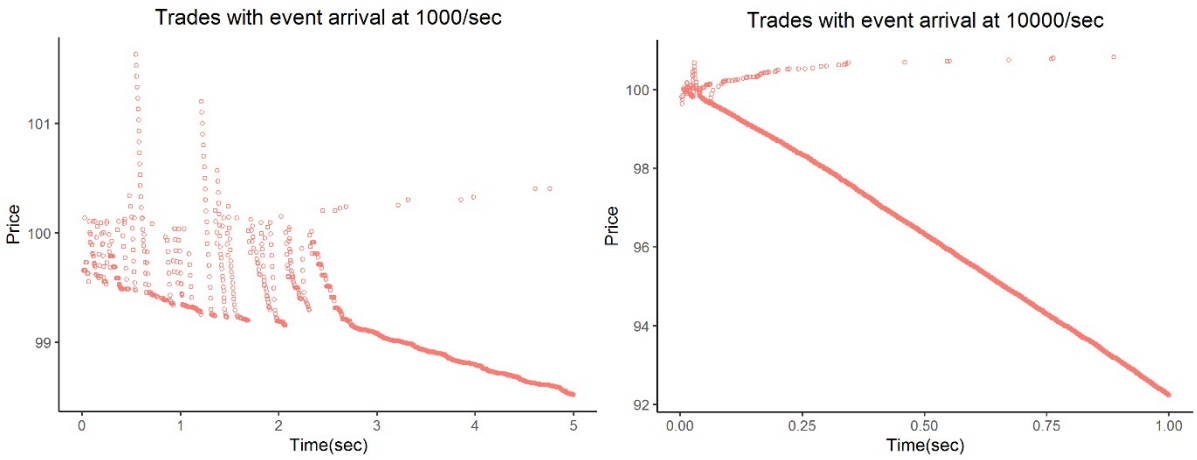


Figure 15: The behaviour of prices under the model. The plots of price and time provides insight into the behaviour of prices. All samples are for trade prices were obtained in the numerical simulation under an environment of bias, i.e traders can quote different quantities. The parameter $\gamma = 1{:}50$, signifying superdiffusive search regime. The plots represent various rates of event arrival - 10, 25, 50,100,1000, 10000 per second. The event arrivals include order arrivals and a driver of the resource density.

CHAPTER 4

LIQUIDITY IN HIGH RESOLUTION IN LIMIT ORDER MARKETS

This chapter reports an investigation, in high resolution, into the role of liquidity in a limit order market. A few dimensions of liquidity are directly studied instead of using the price series as a proxy (an indirect method used by several studies). The implicit assumption in studies that use prices as a proxy for liquidity is that markets are informationally efficient or there is perfect competition. Pani (2020) argued that this assumption may not be as straight forward. Pani (2020) and Hasbrouck (2019) showed that in high resolution assumptions of equilibrium should be cautiously dealt with. To implement this business time is used instead of calendar time and wealth traded in business time is chosen as a proxy for the transaction prices.

Does liquidity in high resolution have any explanatory power on the evolving transaction prices in limit order markets? This question is explored in the chapter. In the proposed model, liquidity in these markets has two parts - a liquidity store and liquidity flow (time dimension). Liquidity in the limit order market is modelled as a liquidity store and incoming information from the

latent order book, to add or cancel liquidity, is modelled as a liquidity flow. A fully probabilistic model is employed. The dimensions covered are depth (five levels in the order book, each on buy and sell side), rate of change in this liquidity on bid and ask side (through order placement and cancellation) and the rate of wealth traded through the transactions. This model includes the concept of Total order book (Limit order book plus Latent order book), a notion that has found acceptance with researchers.

Using a sample of stocks from NASDAQ 100 it is found that liquidity residing in the order book, within five levels each on either side of the transaction price and the flow to this liquidity pool from the latent order book can explain the wealth traded through the trading system, when observed in business time. The explanatory power is satisfactory in high resolution and short periods of business time but weaker in aggregated longer time periods.

The results show, that for the sample of stocks from NASDAQ 100, the broad liquidity can explain the wealth traded through the trading system. The interval score is within acceptable measure of good fit in 54.8 percent of the regimes. A regime constitutes of 300 business time units, where a business time unit is defined as the trade of 100 shares. The explanatory power of liquidity is satisfactory in short periods of business time. The model fit is low in longer periods of time, which suggests that the precision of the signal is low. The conditional nature of the order book and the cancellation behaviour are possible reasons for the low precision. In contrast to the measure of broad liquidity, the liquidity described by the trade clustering, liquidity of marketable orders or liquidity at best bid and best ask prices may be considered as narrow liquidity.

The liquidity pool for large part of the trading day was observed to go into a steady state equilibrium when viewed in business time. It is proposed that when the rate of change of buy side or sell side liquidity is non-zero it coincides with periods of price discovery auctions. In periods of steady state equilibrium, inflow of buy and sell orders through marketable orders and limit orders (including new buyers and sellers) is coordinated with the transaction rates and cancellation behaviour. Algorithmic trading and high frequency trading technology enables precision liquidity management in high resolution. After the price discovery auctions the traders set liquidity conditions that enable wealth transactions at an optimised transaction rate. This is a

partial evidence for the intraday trading model described in chapter 3, where trading was motivated by a need to transform the state of the asset held with the trader.

The liquidity pool is not in steady state during early trading hours and periods of price volatility. A probable inference is that these are periods of price discovery auctions and are likely evidence of higher quality of provisioning by market makers (as discussed in Hasbrouck (2019)). This interpretation is in contrast to Brogaard et al. (2019), who study the case of HFTs and, suggest that when volatility increases the price discovery moves from limit orders to market orders. While, we cannot distinguish between HFTs and other traders in our model and data, observation of the liquidity variables across our sample stocks suggests a role of the broader liquidity during the price discovery auctions. This chapter also contributes in methodology as used in market microstructure research. The proposed model is a fully probabilisitic model (bayesian hierarchial model). The model is trained and used to forecast using Vine Copula Regression as discussed in Cooke et al. (2019). In this way it attempts to expand the microstructure toolbox.

While, there could be several ways to study liquidity, the study presented in this chapter contributes as a high resolution analysis. To contrast, in a recent study on liquidity, Conrad and Wahal (2019) investigate the realized spreads and price impact in a large data set for the period 2010-2017 to draw a term structure of liquidity. The key result in the paper is that the price impact of trades in large (small)-capitalisation stocks takes place within 15 (60) seconds. They recommend the horizons for empirical investigation should be not more than 15 seconds for large capitalisation stock and 60 seconds for small capitalisation stocks. A key issue with this approach is that the bids and asks are asynchronous and hence the spread needs to be estimated. In high resolution, the influence of microstructure effects and other errors would be high.

THE MODEL

We consider a market for a single risky asset. The market is organized as a Total order book which includes the observed limit order book and the unobserved orders of traders that would in the future be placed as limit orders in appropriate time or marketable/orders in appropriate conditions. The limit order book deploys a continuous auction mechanism with the usual price priority. The asset price Y is a random variable, $0<Y<\infty, Y \in R$. Our data allows us to observe the limit order book continuously and the total order book only through the limitorders and

market orders placed. i.e when the orders move from the latent order book to the limit order book. We make no assumptions regarding the latent order book.

Construction of variables

The unit of information in markets today is the orders. When the unit of information was trade prices (in earlier markets) the traditional understanding in learning models of market microstructure was: Traders learn from Buy or sell trades, trade size, number of trades, Time between orders, Time between trades etc. This idea is extended to today when the unit of information is orders. An event in the market consists of a tick. Ticks can be considered to carry an information set called 'marks' {Price, Quantity, Time of arrival, Buy/Sell/Modify/Cancel}; the interaction between the marks results into the transaction prices. The ticks include the bids, offers, cancellations and trades. Every tick is a step in the auction and involves a price-time sorting of the order book. The marks can be further extended to create extended marks such as cumulative liquidity (with various definitions), rate of change in liquidity, order and trade durations, price series for bids and offers, mid-quote price series removing staleness or irrelevance to signal etc depending on the aspect of the information we want to study. Apart from the price and quantity, the traders execute their information through the placement of appropriate quotes (limit, market) and their arrival and departure from the market. Marks contain such information.

Unit of time and liquidity

This empirical exercise uses business time instead of calendar time. The unit of business time chosen is 100 shares traded. Liquidity demand and liquidity supply cannot be distinguished without detailed trader level information. Liquidity can be demanded or made available in both limit orders and marketable orders. The dimensions of this liquidity consist of the volume of shares and activity time (the rate at which this volume changes). Change in total volume available /demanded occurs due to arrival or departure of trades and/or traders into the market. Buy and sell limit orders, transactions involving marketable orders and order cancellation are responsible for changes in liquidity.

Variables

The transactions and limit orders placed / resting in defined band around the best bid and best ask. The 5 best price levels from the best bid (ask) (highest bid (lowest Ask)) is assumed as a region where the transaction price has a high probability to emerge during the trading day. The model consists of five variables described below:

- $h10(Dt)$: Liquidity in top 5 levels each on Bid side and Ask side. This is the cumulative volume of shares in all orders residing in the 10 best price levels.
- $RateBidLiquidity(RBL_t)$: Rate of change in bid side liquidity within each business time unit. The change in cumulative bid volume in the 5 best bid price levels (5 highest bid price levels) within adjacent bid side event updates in calendar time gives the RateBidLiquidity. To compute this, the change in the cumulative bid side volume due to a new event (bid order (addition or modification), order cancellation (removal), trade (removal) is computed if it induces change in the liquidity offered in 5 best bid prices. The calendar time from the last bid side event is used to find the rate of bid liquidity changes.
- $RateAskLiquidity(RAL_t)$: Rate of change in Ask side liquidity within each business unit of time. The computation is similar to the method described for RateBidLiquidity.
- $RateWealthTrade(RL_t)$: Rate of trade transactions. Since the business time is defined as trade of 100 shares, the rate of trade transactions is computed using the cumulative volume weighted transaction within a business unit, divided by the calendar time since last business time unit.
- $WealthTraded(WT_t)$: The WealthTraded in each business unit is the volume weighted Price ($PriceWAP$) multiplied by the number of shares that constitute the business unit i.e 100 shares. $WT_t = PriceWAP * 100$ The volume dimension is captured through the variables Depth10 and the time dimension through the variables RateBidLiquidity, RateAskLiquidity for limit order arrivals and cancellations and RateWealthTrade for the marketable orders.

Empirical hypotheses

The primary empirical hypothesis in this paper is that liquidity (and its dimensions as defined in this chapter) can explain returns in wealth transacted. A probabilistic model is built, trained and

then used to forecast using Vine Copula Regression as discussed in Cooke et al. (2019) (CBJ). The idea is that if explanatory variables and a response variable of interest are simultaneously observed, then fitting a joint multivariate density to all variables would enable prediction via conditional distributions. Regular vines or vine copulas with arbitrary univariate margins provide a rich and flexible class of multivariate densities for Gaussian or non-Gaussian dependence structures. The density of the the vine copulas enables calculation of all regression functions. This method helps avoid issues of transformations, heteroscedasticity, interactions, and higher-order terms (CBJ).

The evolution of wealth traded through $PriceWAP$, similar to the evolution of prices is considered and an explanatory model is built by mapping the one period forward wealth traded. The proposed model is the multivariate function:

$$f\left(D_{t-1}, RBL_{t-1}, RAL_{t-1}, RT_{t-1}, WT_t\right) \tag{17}$$

The following conditional distribution from (17) above can be estimated using Vine Copula Regression:

$$f\left(WT_t \vee D_{t-1}, RBL_{t-1}, RAL_{t-1}, RT_{t-1}\right) \tag{18}$$

The estimation is done by implementing the vine copula regression procedure set in Chang and Joe (2019). The steps are given below:

1. Univariate distributions are chosen. Completely probabilistic models (Bayesian / Bayesian hierarchical models) for the univariate distributions are chosen. The distributions range from lognormal $\left(Depth10, WealthTraded\right)$, skew-normal $\left(RateBidLiquidity, RateAskLiquidity\right)$ to exponential $\left(RateWealthTrade\right)$. The details of the marginal models are provided in the chapter. The corresponding u-scores are obtained by applying the probability integral transform.
2. A Vine copula is fit on the u-scores using two steps. First, an appropriate vine copula structure is chosen. The largest absolute partial correlation is taken in the lowest and second, and appropriate bivariate copulas is chosen from a set {1 = Gaussian copula, 2 = Student t copula (t-copula), 3 = Clayton copula, 4 = Gumbel copula, 5= Frank copula, 6 =

Joe copula} using the model selection criteria AIC. Tree 1 of the vine is found with a maximum spanning tree algorithm over a complete graph. The copulas are fitted sequentially from Tree1 onwards as implemented in R package VineCopula.

3. The vine copula gives the joint distribution and the vine copula regression finds the conditional distribution of the response variable (WT_t) given the explanatory variables $(D_{t-1,}, RBL_{t-1}, RAL_{t-1}, RT_{t-1})$.
4. The next step constructs a truncated R-vine that satisfies a constraint. Starting with a locally optimal t-truncated R-vine using the explanatory variables, the response variable is sequentially linked to the node that satisfies the proximity condition and has the largest absolute correlation with the response variable. The data structure has 4 explanatory variables and one response variable arranged as $f(X, Y)$. Based on the observation in Chakravarty and Pani (2018) that the R-Vine copula structure can change over time during a trading day, the trading day is divided into regimes for estimation. Equal sized regimes of 300 data points are taken. Every regime is a sample to fit a flexible parametric vine copula and the succeeding regime is the out-of-sample (oos) observation to test the predicted values from our model. The choice of the number of data points in each regime gives sufficient number of data points to enable estimation, but avoids over fitting that can destroy information in high resolution.

Prediction

Model performance is validated using out-of-sample (oos) data set. The fitted vine copula regression model is used to predict the conditional distribution of the *WealthTraded* using the oos data-set. First the conditional CDF of the response variable is computed and then the conditional quantile and mean is calculated by inverting the conditional CDF and integrating the quantile function. The method as implemented in Chang and Joe (2019) is used. The performance is evaluated by applying the root-mean-square error (RMSE), predicted log-likelihood (loglik), Interval score (IS) and quadratic score (QS) (Gneiting and Raftery (2007)).

To repeat the process, a regime based moving window approach with no overlap between testing regime is implemented to keep the estimation sensitive to changing dynamics within the

variables and between the variables. Sequentially, the oos data sample becomes the training dataset for next regime.

RESULTS AND DISCUSSIONS

The empirical estimation is based on data sample from NASDAQ Itch. The data comprises a random sample of 41 stocks from NASDAQ 100 index for the trading day 27-Mar-2019. The data is first prepared based on the data handling paradigm proposed in chapter 6. This data records all the messages - orders placed (buy, sell, modify, cancel) and trades executed in the exchange. order and message ids in the data can be used to model and recreate the limit order book. The timestamp is in nanoseconds and can be used for any high resolution analysis.

The empirical model explores the role of liquidity that resides in the five best buy and sell price levels around the transaction price. Marketable buy (sell) orders from the Total order book can climb up these price levels to transact with limit sell (buy) orders.

Such orders are not directly observable but indirectly inferred using the transaction rate. Liquidity in *Depth* 10 gets built up with placement of incoming buy and sell orders. At the same time it gets eroded with cancellation and transactions. These changes are tracked by computing the rate of change in buy side liquidity, sell side liquidity and transaction rate.

The data and variables of a sample stock, AMAT, for the trading day 27-Mar-2019 is described, followed by the performance of the model for this example. Then the results of the full sample and performance of the model is reported.

The description of the trading day for a sample stock (AMAT) shows that the random observation is anecdotally in line with the model in Chapter 3. That model looks at the trading system as a means for traders to achieve a change in state of their holdings (cash to shares and vice versa). The trading system is seen as learning and observing the information content of quotes (orders and trades), with information absorption taking place into the low frequency public signal through the pre-auctions and auctions in early trade. Intraday prices and the public signal are part of the price system. Intraday prices can get generated based on the ‘auction as a search’ in that model.

Steady state in liquidity conditions

For large expanses in business time, the liquidity conditions remain in steady state conditions. This may be defined as conditions in which the rate of change in buy side liquidity and/or sell side liquidity is zero. The total liquidity remains constant. Based on the above observations the intuition is the following:

When the rate of change of buy side or sell side liquidity is non-zero, it coincides with periods of price discovery auctions. Note that price discovery does not necessarily imply price changes. As shown in figure 16 prices change even in the steady state liquidity conditions. Trading system participants are not perturbed if price changes are within expected volatility limits. Liquidity conditions is not static when these two rates goes to zero. We know that the transaction rate is non-zero based on our definition of business time and as seen in figure. 17. There is also exit of liquidity through the cancellation behaviour. Hence, inflow of buy and sell orders through marketable orders and limit orders (including new buyers and sellers) is coordinated with the transaction rates and cancellation behaviour. Algorithmic trading and high frequency trading technology enables precision liquidity management in high resolution. Once the traders negotiate the price discovery auctions (including periods of price impact of large trades) when surprises in new prices can emerge, they account for changes in prices due to the volatility and set liquidity conditions that enable wealth transactions at an optimised transaction rate (similar to the state transformation model described in Chapter 3).

Our intuition and conjecture in the preceding paragraph should be viewed in conjunction with findings in Hasbrouke(2019) and merits further clarification. The stock in question (AMAT) trades in a fragmented market with primary listing in NASDAQ. VAR and VECM techniques applied to the transaction prices may not support our proposition on price discovery auctions. However, as shown by Hasbrouke(2019) in high resolution the information share of exchange where the security is listed is substantially higher than its share in trading. While the paper reasons the higher information share to the presence of designated market makers, we provide evidence of what may constitute a higher quality of liquidity provisioning from such participants.

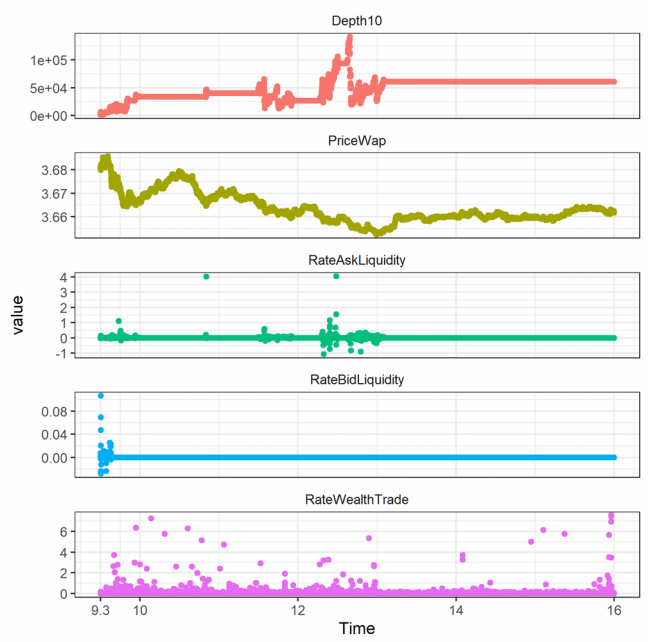


Figure 16: Plot describing the variables in calendar time for AMAT (NASDAQ), trading day 27-Mar-2019, 9.30 am to 4.00 pm. The variables Depth10, RateBidLiquidity and RateAskLiquidity go into steady state equilibrium or inactivity during parts of the trading day.

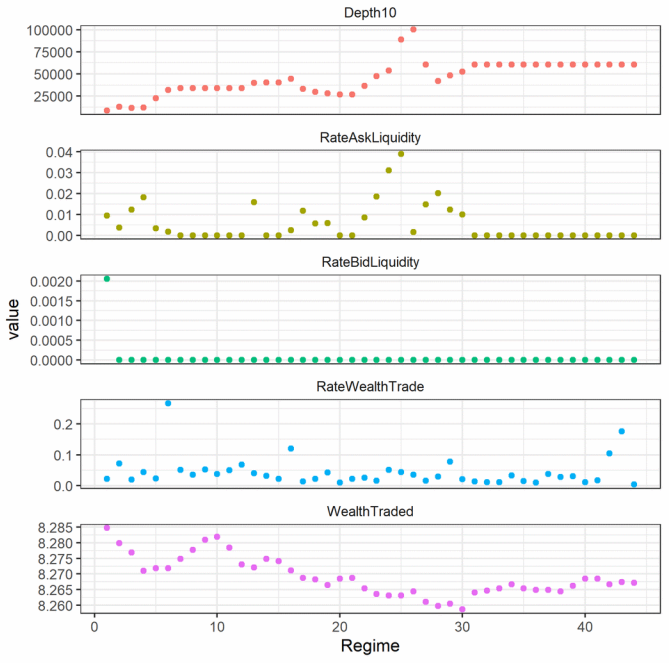


Figure 17 Plot describing the variables in business time for AMAT (NASDAQ), trading day 27-Mar-2019, 9.30 am to 4.00 pm. The variables Depth10, RateBidLiquidity and RateAskLiquidity are aggregated at the regime level. Each regime consists of 300 business time units. A business time unit comprises transaction of 100 shares. There are 44 regimes for this stock in the trading day. Behaviour of the variables is visually lagged in business time showing higher amount of trading activity in first half of the day.

Performance of the model: Example of AMAT (NASDAQ)

The model is evaluated using the predictive performance on out of sample observations and the results are reported. The oos log-likelihood, Interval score and quadratic score are relevant for probabilistic models. We report the estimation results for 41 regimes. The oos log-likelihood average is -6.7214. The range is between -20.5869 to -2.8014. The log-likelihood was not available in 15 of the 41 regimes. The interval score however is available for all regimes. The average interval score is 8.462. The interval scores of the individual regimes is more encouraging.

Around 50% of the interval scores is below 5 as seen in figure 18. Such an interval score demonstrates good performance of the model in oos samples. Figure 19 plots the actual versus the predicted log-WealthTraded. This plot challenges the notion of a good fit of the model and underscores the importance of scoring rules used by us to evaluate performance. A closer visual fit is more important in deterministic models. The scoring rules more than adequately test the performance of our probabilistic model. We do not expect the liquidity model to explain the evolution of the transaction price or PriceWAP or the wealth traded completely.

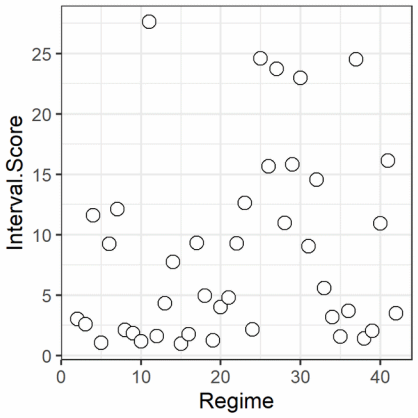


Figure 18 Interval Scores for AMAT (NASDAQ, trading day 27-Mar-2019), depicting the out of sample predictive performance of the model in all the regimes. The average interval score is 8.462. Around 51% of the interval scores in individual regimes is below 5 showing good model fit.

There exists no temporal or regime pattern to the better fit in the models. The average quadratic score for all the regimes is 0.0255. The distance of the predicted probability distribution from the true probability distribution is high. The highest quadratic score of 0.0991 is obtained in the 24th regime. There are five other regimes with quadratic scores greater than 0.08.

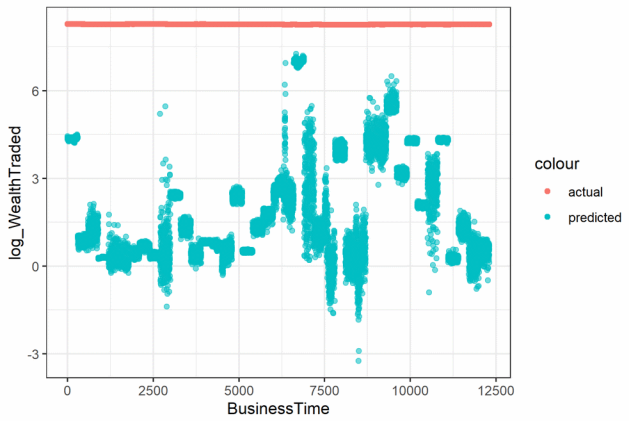


Figure 19: Wealth traded (log scale): Actual versus Predicted in business time (AMAT (NASDAQ) trading day 27-Mar-2019)

Full Sample: Description of data

The full sample comprises of 41 stocks randomly selected from the NASDAQ 100 and trading in NASDAQ on 27-March-2019. The variables are computed at the business time intervals. A business time unit is the trade of 100 shares.

Calendar time has also been used where ever required, for example, to compute the rate variables. The estimation is based on (18) and was carried out independently in regimes consisting of 300 business time units or 300 data points each. There are a total of 1283 regimes and a total of 1242 regimes for out of sample testing. The estimation in each regime was then evaluated in out of sample data from the following regime. The results of this performance evaluation has been presented and discussed. In this section the distribution of the liquidity variables is described. First the means of the variables for each stock is computed.

Figure 20 provides the distribution of the liquidity variables for our sample. The Depth10 has a few large outliers resulting in a fat tail to the right (a log transformation of Depth10 is used in fig

18 for illustration). The average Depth10 is 114180 while the median is only 1818. The max Depth10 in the sample is 3144752. The very large values is not necessarily from the popular stocks or the large stocks where maximum trade occurs. The distribution of RateWealthTrade visually looks similar but the illustration is not on the log scale. The mean (0.1990) and median (0.1119) are close. There does exist a smaller secondary peak in the density. Overall for the huge range in Depth10 the range in RateWealthTrade is narrow (Max (1.6031)-Min(0.0166)).

The effect gets further dampened when the WealthTraded (output of the trading system) is observed. The mean (9.2087) and median (9.2531) are further close. The range in WealthTraded is Max (11.6701)-Min(6.3431). This variable is in the log scale. The density plots of RateBidLiquidity and RateAskLiquidity are provided after removal of 2-3 outliers. These plots are based on the averages of the variables at the individual stock level. The density plots for both the variables show high kurtosis with their medians, -5.92E-07 and -7.311E-05 respectively, being close to zero. This explains the vast expanses of business time when these variables are close to zero. The RateAskLiquidity is more distributed than RateBidLiquidity and also has more secondary smaller peaks. The range for RateBidLiquidity is max (0.2356) to min(-0.0055). Similarly, the range for RateAskLiquidity is max (0.016) to min (-1.476).

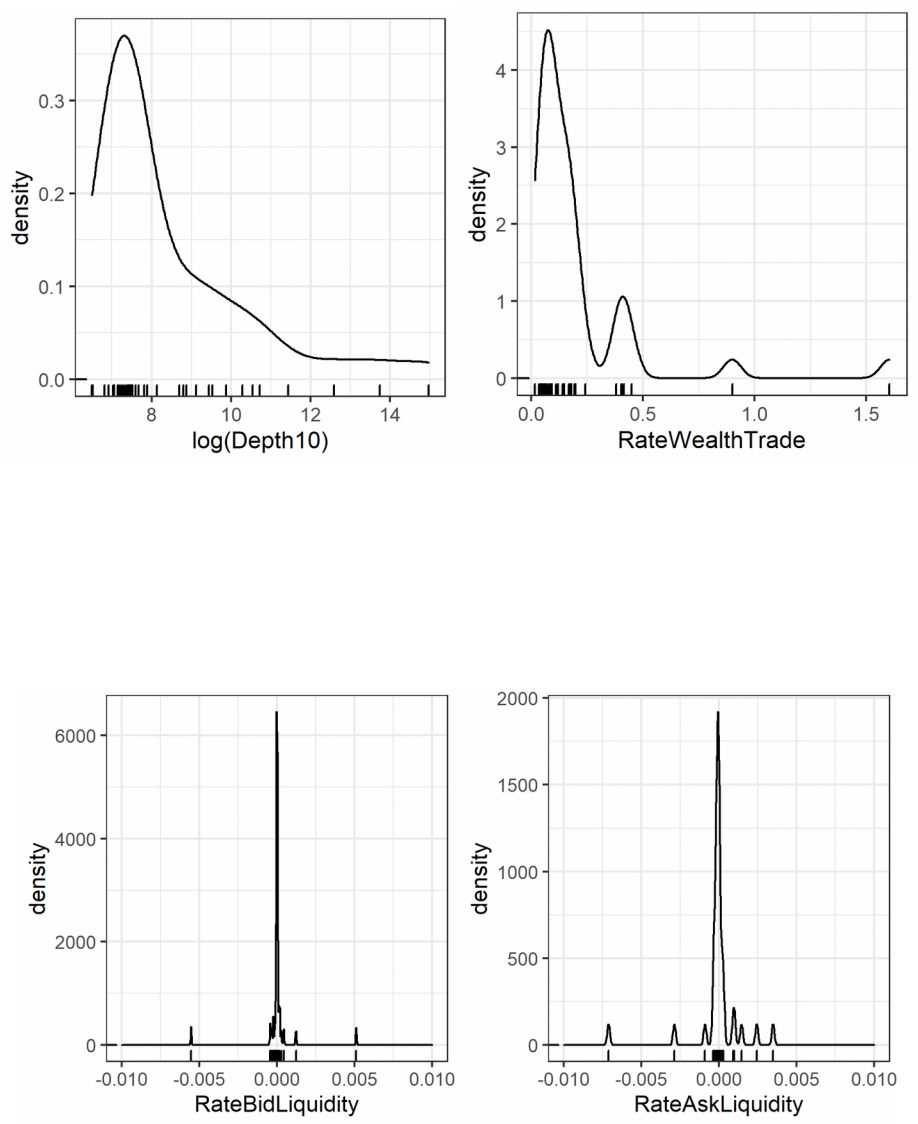


Figure 20: Description of the liquidity variables for the sample of 41 stocks. The variables described through the density plots are Depth10 (log transformed due to outliers), RateWealthTrade, RateBidLiquidity and RateAskLiquidity.

Liquidity store and liquidity flow

Liquidity variables are of two kinds - liquidity store (Depth10) and liquidity flow (RateBidLiquidity, RateAskLiquidity, RateWealthTrade). The entire liquidity store (Depth10) available at any instant in the trading system may not eventually be traded. Under the assumption that the liquidity store would actually be completely utilised, Depth10 can be expressed in terms of business time units. This is expressed as *LiquidityAvailable*. The liquidity traded in terms of

business time is *LiquidityTraded*. The average store (*LiquidityStore*) can be expressed as a percentage, *LiquidityAvailable* to *LiquidityTraded*. For AMAT this is 3.59%. The distribution for LiquidityStore for our sample is provided in figure 21. 22 stocks (53.65%) have LiquidityStore less than 1 percent. Another 9 stocks (22%) have LiquidityStore between 1-2.5 percent. 3 stocks (7.3%) between 2.5-5 percent and 2 stocks (5%) between 5-10 percent. 12% of stocks are outliers holding LiquidityStore beyond 10 percent. We do not find any trends or patterns in this and the distribution seems to be related to the individual microstructure of the stocks.

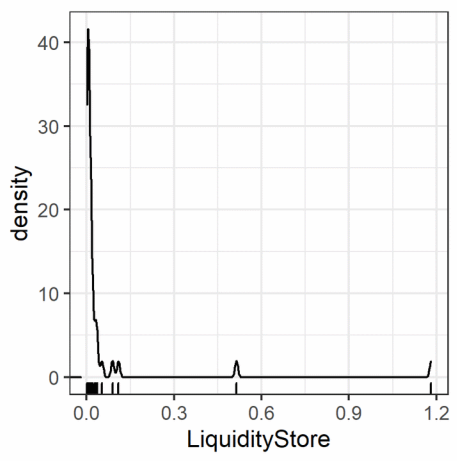


Figure 21: Distribution of LiquidityStore. LiquidityStore (expressed as a percent) is the amount of liquidity available in average Depth10 compared to total liquidity traded. This is computed by first arriving at LiquidityAvailable and LiquidityTraded in terms of business time units. 75% of the sample has LiquidityStore less than 2.5% .

Full Sample: Performance of the model

This section reports the performance of the model in out of sample data. There is a total of 1242 out of sample regimes across 41 stocks where the model predictions were tested. To evaluate the performance the log-likelihood, the interval score and the quadratic score were computed. The log-likelihood exists in 50.4 percent or 626 forecasted regimes. At individual stock level the range is 0 (CHKP) -100 (VRSK, CTAS, ISRG) percent. The average log likelihood is -6.9781. The range among averages at stock level is -19.622 (COST) to -3.5729 (MELI). The interval score has been calculated for all the regimes. The acceptable measure of good fit is an interval score of less than 5. The overall average of interval score is 6.8637. The range for the average interval score among the individual stocks is 0.8025 (VRSK) to 15.677 (MELI). In 10 stocks (25%) the interval score average at the individual stock level is less than 5. Since the estimation

and forecasts has been done independently at regime level, the performance at the regime level can be evaluated. In a total of 680 regimes or 54.8 percent the interval score is less than 5. This regime level model fit is found to be encouraging. The range for the percentage of regimes at stock level, where the interval score is less than 5, is 25 (GOOG) to 100 (VRSK, HAS, CTAS, ISRG, CHKP) percent. The regimes where the log likelihood did not exist, returned a zero quadratic score. In these nearly 50 percent of the regimes the distance of the forecasted distribution from the true distribution is high. For the remaining 50% there exists some level of model fit. The overall quadratic score for the sample is 0.0267. The range for the average quadratic score at the stock level is from 0 (VRSK) to 0.0803 (SIRI).

Depth10: Information and Precision

The liquidity model attempts to explain the wealth traded at any business time using the one period lagged liquidity variables Depth10, the change of this liquidity on the bid side and ask side as well as the change in the rate of trade. Depth10 is a notion of broad liquidity. A volume of demand or supply at a set of possible price points that may result into transaction. This provides the information or prediction potential to Depth10. Some of the orders constituting Depth10 may never get transacted before getting cancelled. Thus, lower precision in the signal is expected compared to measures based on trades or marketable orders or the best bid and best ask. Further, Depth10 cannot completely define liquidity even at broader level and is associated with the time dimension through the flow of liquidity. The time dimension is captured through the rate variables. Since the transaction rates (RateWealthTrade) get linked to liquidity available in Depth10 we expect the existence of some contemporaneous and predictive ability of liquidity to predict one period ahead WealthTraded. The predictive performance reported shows that the liquidity model is relevant at the regime level and has reasonable fit for a probabilistic model. However, there exists a mixed signal coming from the overall model as depicted by the overall sample and stock averages of interval score and the absence of log likelihood and quadratic score for 50 % of the regimes. This points to the expected lower precision from the model.

A key feature of Depth10, noted earlier, was the transition to steady state characteristics during the trading day. Such a feature is to be expected in markets where the dominant traders are institutional and equipped with fast trading infrastructure and also algorithmic trading

capabilities. 5 price levels on either side of the transaction price is an arbitrary choice based on intuition and observation with a few exemplar stocks. The idea is that there can exist broader notions of liquidity other than the transaction clustering or supply, demand and imbalances of supply and demand at trades, marketable orders and best bid and best ask levels. This idea has been conceptually and empirically demonstrated in this chapter.

The Rate variables

The rate variables RateWealthTrade, RateBidLiquidity and RateAskLiquidity complete the liquidity model when added to Depth10. Independently, the rate variables do not offer much insights. The RateWealthTrade is consistently present throughout the trading day as a result of the design where the business time unit is defined as transaction of 100 shares. However, RateBidLiquidity and RateAskLiquidity are characterised by short spells of activity and then transition to near zero values. Such spells combine with the steady state characteristics of Depth10. The interpretation proposed for these periods of activity of the two key rate variables is that they represent Price Discovery auctions. Such auctions exist during the first trading periods of the trading day and more volatile periods where there has been volatile price movements. It is suggested that designated market makers and Institutional traders in these counters step up their search for the acceptable price. Once the price is settled, they can go back to steady state allowing machines to execute a change in state of their holdings. Such an inference is in contrast to Brogaard et al. (2019) who suggest that when volatility increases the price discovery moves from limit orders to market orders in case of HFTs. Although specific behaviour of HFT cannot be distinguished in the proposed model, observation of the liquidity variables across a few sample stocks suggests a role of the broader liquidity during the price discovery auctions.

CHAPTER 5

A DATA PARADIGM TO HANDLE HIGH RESOLUTION ORDERS AND TRADES DATA

This chapter, a joint work with Dr. Ranjan Chakravarty, presents a data paradigm to handle high frequency data for empirical research in high resolution. It reports discovery and observation of

quotes and trades that do not correspond to plausible market activity as in Brownlees and Gallo (2006).

Market microstructure data consists of trade and quotes, signals and outliers. It is observed that in UHF data white noise emanates not only from outliers but also from specific stochastic processes. The presence of such stochastic processes (quotes and trades) is an impediment to operationalize the use of expanded filtration in empirical research.

Existing data handling paradigm such as Brownlees and Gallo (2006) and Barndorff-Nielsen et al. (2009) can handle data emerging from trades and the best bids and offers. In order to include the remaining orders into the datasets, the proposed data paradigm relies on the theoretical extension provided in Kchia and Protter (2015) (extended by Neufcourt and Protter (2019)) and proposes to operationalise the filtration expansion established in that paper. The efficacy of the proposed data paradigm is presented in figure 22.

It specifies, by invoking Brownlees and Gallo (2006), that in auction markets quotes that participate in the auction in search of a trade are considered to be plausible market activity. Quotes that do not comprise plausible market activity are termed non-microstructure noise. Identifying and removing those quotes that do not represent plausible market activity are seen to be key to improved estimation in practice.

This chapter makes three specific contributions to the existing empirical microstructure literature.

- First, it shows that in limit order markets there exist stochastic noise processes in addition to and different from microstructure noise. This discovery in effect throws a challenge to use filtration expansion techniques in empirical estimation in high resolution.
- Second, it proposes a new paradigm to identify these noise processes and remove them. This proposed paradigm paves the way to use filtration expansion techniques.
- Third, it empirically demonstrates how the application of this paradigm results in vastly improved parametric estimation.

An empirical example has been included to demonstrate the efficacy of the data paradigm. Along with this the empirical example also suggests a novel way to estimate the precision of

the signal in the incoming quotes. The Realised kernel (based on Barndorff-Nielsen et al (2009)) is estimated from the mid-price series constructed from incoming quotes and also from transactions. The construction of the mid-price series is novel, as it attempts to prevent staleness of the data series and at the same time not lose track of the evolving price. It is constructed by considering the mid-price of say an incoming bid with the existing best ask (and vice-versa).

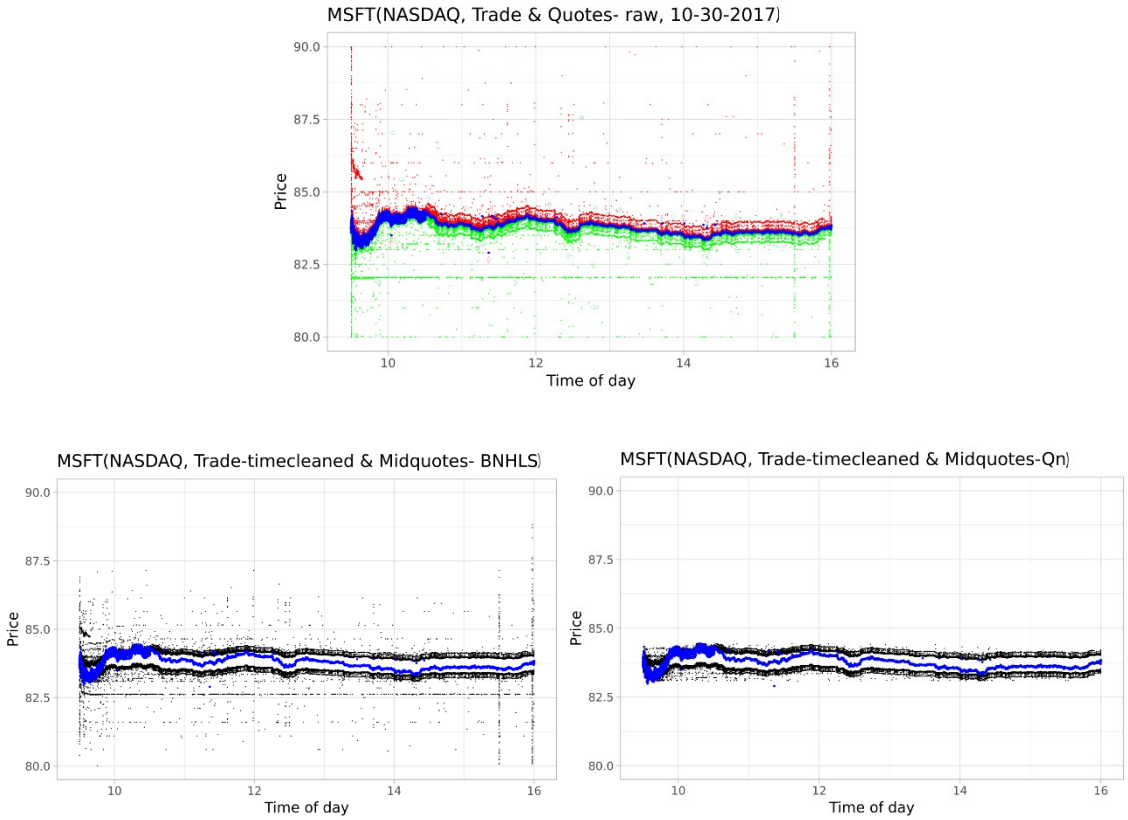


Figure 22: Comparative efficacy of the new data paradigm with the existing Barndorff-Nielsen et. al (2009) (BNHLS). The raw data is presented in the top panel. Bids (green), Asks (red) and trades (blue). The bottom panel shows trades in blue as above and the cleaned mid-quote price series. The right hand side of the bottom panel which shows why the new data paradigm provides better estimation.

In this way with every data point incoming information into the trading system is incorporated. The effect of data cleaning is to improve estimation. Hansen and Lunde (2006) had made intuitive comments regarding this. The interpretation of the Realised kernel estimated is as follows: First, our data cleaning using lob modelling and other methods based on the central tendency, have ensured that every data point joining the midquote is part of the expanded

filtration. Second, the realised kernel of the midquote estimates the maximum information coming into the market under the assumption that each incoming quote had an opportunity to participate in the auction and get transacted. Third, the realised kernel from the transaction prices estimates the true signal coming from the prices. It includes the information already incorporated into the asset price.

CHAPTER 6

CONCLUSION, LIMITATIONS AND FUTURE RESEARCH OPPORTUNITIES

Conclusion

This dissertation has explored research topics in the area termed as high-resolution market microstructure, an emerging and active research area. The introduction of electronic trading platforms in the late 1990s led to the democratisation of stock exchanges. Orders from participants were received directly at the exchange and not an intermediary. It also led to the emergence of algorithmic trading. Both these developments resulted in orders being the unit of information in limit order markets. Further, with improvements in technology-driven trading infrastructure, each step in the trading process became automated and faster.

Market microstructure models lagged in incorporating changes. An 'auctioneer' or 'market maker' is introduced in most models, although designated market makers do not exist now, unlike in the 1980s and 1990s. This design was essential to assume the emergence of prices from an equilibrium of supply and demand. A research gap that the essays in this thesis attempt to fill in. The microstructure toolbox was not sufficient to the needs of using orders as the unit of information.

The intraday trading model, proposed in chapter 2, models a limit order market using a total order book and the absence of market makers. The proposed model is closer to describing current order-driven markets. The model accepts the emergence of prices from an alternate tatonnement mechanism – 'auction as a search' introduced in chapter 3. The tatonnement mechanism does not arbitrarily assume the existence of an equilibrium in high resolution. It is a useful addition to the microstructure toolbox for high-resolution studies.

Existing empirical studies in high resolution have provided two important empirical results. The first uses transaction prices to demonstrate the nature of price discovery in high resolution. The second shows that price discovery also existed in limit orders of high-frequency traders. It uses the orders resting in the first five price levels of the limit order book for the research. Chapter 4 has added to this by showing that the liquidity provided by the first five levels in the order book (both buy and sell side) has explanatory power over the returns in wealth traded (proxy to the transaction prices) through the trading system. This explanation is satisfactory in short periods of business time. To establish this role of liquidity in high resolution, chapter 4 uses probabilistic models, pair copula construction and vine copula regression. The essay in chapter 4 may be the first applications of these methods in market microstructure. It could serve as a useful addition to the microstructure toolbox.

However, before the empirical investigation in chapter 3, data handling challenges had to be addressed. The empirical investigation of the information contained in orders is challenging. Firstly, the theoretical justification of connecting the orders to transaction prices. Secondly, the discovery of new stochastic processes termed non-microstructure noise. Chapter 5 reviews and presents the required theoretical justification that existed outside the boundaries of market microstructure. A new data paradigm that handles ultra-high-frequency data for studies using the orders and trades data is presented. The key intuition was to determine if an order participated in the continuous auction and whether the trade transaction was representative of the continuous auctions. This research further adds to the existing data handling techniques available in the literature and highly relevant addition to the microstructure toolbox when considering high-resolution studies.

Limitations and future research opportunities

The last section of the dissertation lists the limitations of this research and future research opportunites.

A few of them are presented in this synopsis.

The intraday trading model assumes that the public signal remains constant during the day. This assumption may seem arbitrary and was used to simplify the model. One benefit of the current design is that the public signal updates through a pre-auction mechanism. Update through a pre-

auction mechanism also supports notions such as the information content structure. Any different assumption for this variable should also incorporate a possible information aggregation mechanism. Other possible assumptions may lead to different implications.

While formal or designated market makers are not present in a large number of stock exchanges, contributions from market-making activities voluntarily provided by institutional traders (including algorithmic or high frequency traders) have not been considered. Such a contribution exists and is well documented. Improvements in the intraday trading model may consider this in future.

The continuous double auction mechanism considered in this research is the popular design with price-time priority. However, competitive forces have ensured that in a fragmented market, exchanges offer different design to attract trading volume. Different kind of traders may prefer a mechanism other than the price-time mechanism. Mathematical treatment and outcomes would vary when this change is effected. We have considered the most widely used mechanism, and the current design did not have the scope to consider alternatives.

In the empirical study on the role of liquidity in high resolution, the empirical model used a mix of business time and calendar time. Business time is used while estimation and updation to the variables continued with each tick at calendar time. This design has its own advantages. However, one may consider a design only with calendar time to check on the implications. Specifically, the transactions would not get homogenised as it does when business time transformations place. The non-synchronous nature of the data at high resolution presented difficulties for the researcher during estimation.

The equity markets in the US and Europe are fragmented. The same asset is traded simultaneously in different markets and sometimes in different time zones in either overlapping and non-overlapping manner. Fragmented markets is also an active area of research in itself. However, the current research could not incorporate the same in the design.

Depending on the research questions involved, future research may consider the incorporation of multiple venues of trading. The current research design also does not invoke a complete market – incorporating the asset and derivatives. Although most microstructure research design do not consider complete markets, it may become imperative when one considers the reality in certain

emerging asian markets which have very robust and extensive trading of stock and index derivatives.

REFERENCES

A. I. McLeod and W. K. Li. Diagnostic Checking ARMA Time Series Models Using Squared-Residual Autocorrelations. Journal of Time Series Analysis, 4(4):269–273, July 1983.

Abhishek Sanskrityayn and Naveen Kumar. Analytical solution of advection diffusion equation in heterogeneous infinite medium using Greens function method. Journal of Earth System Science, 125(8):1713–1723, December 2016.

Albert J Menkveld. The economics of high-frequency trading: Taking stock. Annual Review of Financial Economics, 8:1–24, 2016.

Albert S Kyle, Anna A Obizhaeva, and Yajun Wang. Smooth trading with overconfidence and market power. The Review of Economic Studies, 85(1):611–662, 2018.

Albert S Kyle. Continuous auctions and insider trading. Econometrica: Journal of the Econometric Society, pages 1315–1335, 1985.

Alfonso Dufour and Robert F. Engle. Time and the Price Impact of a Trade. The Journal of Finance, 55(6):2467–2498, December 2000. ISSN 00221082.

Anat R. Admati and Paul Pfleiderer. A Theory of Intraday Patterns: Volume and Price Variability. Review of Financial Studies, 1(1):3–40, January 1988. ISSN 0893-9454, 1465-7368.

Andrei Kirilenko, Albert S. Kyle, Mehrdad Samadi, and Tugkan Tuzun. The Flash Crash:

Andy M. Reynolds. Current status and future directions of Levy walk research. Biology Open, 7(1), January 2018.

Avraham Beja and M Barry Goldman. On the dynamic behavior of prices in disequilibrium. The Journal of Finance, 35(2):235–248, 1980.

Avraham Beja and Nils H Hakansson. Dynamic market processes and the rewards to upto date information. The Journal of Finance, 32(2):291–304, 1977.

Bo Chang and Harry Joe. Prediction based on conditional distributions of vine copulas. Computational Statistics & Data Analysis, 139:45–63, 2019.

Bruno Biais, Fany Declerck, and Sophie Moinas. Who supplies liquidity, how and when? TSE Working Papers 17-818, Toulouse School of Economics (TSE), June 2017.

C Godreche and JM Luck. Statistics of the occupation time of renewal processes. Journal of Statistical Physics, 104(3-4):489–524, 2001.

C.T. Brownlees and G.M. Gallo. Financial econometric analysis at ultra-high frequency: Data handling concerns. Computational Statistics & Data Analysis, 51(4):2232–2245, December 2006.

Charles R Plott. Properties of disequilibrium adjustment in double auction markets. Handbook of experimental economics results, 1:16–21, 2008.

Charles-Albert Lehalle and Othmane Mounjid. Limit Order Strategic Placement with Adverse Selection Risk and the Role of Latency. Market Microstructure and Liquidity, 03(01):1750009, March 2017. ISSN 2382-6266, 2424-8037.

Christine A Parlour and Duane J Seppi. Limit order markets: A survey. Handbook of financial intermediation and banking, 5:63–95, 2008.

D. Froemberg, M. Schmiedeberg, E. Barkai, and V. Zaburdaev. Asymptotic densities of ballistic Levy walks. Physical Review E, 91(2), February 2015.

Daniel Grunbaum. Using Spatially Explicit Models to Characterize Foraging Performance in Heterogeneous Landscapes. The American Naturalist, 151(2):97–113, February 1998.

David Easley and Maureen O'Hara. Price, trade size, and information in securities markets. Journal of Financial Economics, 19(1):69–90, September 1987. ISSN 0304405X.

David Easley and Maureen O'hara. Time and the process of security price adjustment. The Journal of finance, 47(2):577–605, 1992.

David Easley and Maureen O'Hara. Time and the Process of Security Price Adjustment. The Journal of Finance, 47(2):577–605, June 1992. ISSN 00221082.

David Easley, Marcos Lopez de Prado, and Maureen O'Hara. Discerning information from trade data. Journal of Financial Economics, 120(2):269–285, May 2016. ISSN: 0304405X.

Denis Gromb and Dimitri Vayanos. Equilibrium and welfare in markets with financially constrained arbitrageurs. Journal of Financial Economics, 66(2-3):361–407, November 2002. ISSN 0304405X.

Dimitri Vayanos et al. A dynamic model of an imperfectly competitive bid-ask market. Technical report, European Science Foundation Network in Financial Markets, c/o CEPR, 33 Great, 1993.

Dimitri Vayanos. Strategic trading and welfare in a dynamic market. The Review of Economic Studies, 66(2):219–254, 1999.

Douglas W. Diamond and Robert E. Verrecchia. Constraints on short-selling and asset price adjustment to private information. Journal of Financial Economics, 18(2):277–311, June 1987. ISSN 0304405X.

Duane J. Seppi. Liquidity Provision with Limit Orders and a Strategic Specialist. Review of Financial Studies, 10(1):103–150, January 1997. ISSN 0893-9454, 1465-7368.

E P Raposo, S V Buldyrev, M G E da Luz, G M Viswanathan, and H E Stanley. Levy flights and random searches. Journal of Physics A: Mathematical and Theoretical, 42 (43):434003, October 2009.

Esen Onur, John S. Roberts, and Tugkan Tuzun. Trader Positions and Marketwide Liquidity Demand. Finance and Economics Discussion Series, 2017(103), October 2017. ISSN 19362854.

F. Douglas Foster and S. Viswanathan. A Theory of the Interday Variations in Volume, Variance, and Trading Costs in Securities Markets. Review of Financial Studies, 3(4): 593–624, October 1990. ISSN 0893-9454, 1465-7368.

Fischer Black. Toward a fully automated stock exchange, part i. Financial Analysts Journal, 27(4):28–35, 1971.

Florian Simatos. Coupling limit order books and branching random walks. Journal of Applied Probability, 51(3):625–639, 2014.

Frank R. Hampel. The Influence Curve and Its Role in Robust Estimation. Journal of the American Statistical Association, 69(346):383, June 1974.

Giovanni Cespa and Xavier Vives. High frequency trading and fragility. ECB Working Paper No. 2020, 2017. URL .

High-Frequency Trading in an Electronic Market: The Flash Crash. The Journal of Finance, 72(3):967–998, June 2017. ISSN 00221082.

Ilan Kremer. Information aggregation in common value auctions. Econometrica, 70(4):1675–1682, 2002.

J Donier and J-P Bouchaud. FromWalras auctioneer to continuous time double auctions: a general dynamic theory of supply and demand. Journal of Statistical Mechanics: Theory and Experiment, 2016(12):123406, December 2016.

J. Donier, J. Bonart, I. Mastromatteo, and J.-P. Bouchaud. A fully consistent, minimal model for non-linear market impact. Quantitative Finance, 15(7):1109–1121, July 2015.

Jean Jacod, Yingying Li, and Xinghua Zheng. Statistical properties of microstructure noise. Econometrica, 85(4):1133–1174, 2017.

Jennifer Conrad and Sunil Wahal. The term structure of liquidity provision. Journal of Financial Economics, 2019.

Joachim Grammig, Erik Theissen, and OliverWuensche. Time and price impact of a trade: A structural approach. 2011.

Joel Hasbrouck. Measuring the Information Content of Stock Trades. The Journal of Finance, 46(1):179–207, March 1991. ISSN 00221082.

Joel Hasbrouck. Price Discovery in High Resolution*. Journal of Financial Econometrics, 09 2019. ISSN 1479-8409. doi: 10.1093/jjfinec/nbz027.

John Bone. Risk-sharing cara individuals are collectively eu. Economics Letters, 58(3):311–317, 1998.

Jonathan Brogaard, Terrence Hendershott, and Ryan Riordan. Price discovery without trading: Evidence from limit orders. The Journal of Finance, 74(4):1621–1658, 2019.

Kerry Back, C Henry Cao, and Gregory A Willard. Imperfect competition among informed traders. The journal of finance, 55(5):2117–2155, 2000.

Kerry Back. Insider trading in continuous time. The Review of Financial Studies, 5(3):387–409, 1992.

Larry Harris. Trading and exchanges: Market microstructure for practitioners. OUP USA, 2003.

Lawrence R Glosten and Paul R Milgrom. Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. Journal of financial economics, 14(1): 71–100, 1985.

Léo Neufcourt and Philip Protter. Expansion of a filtration with a stochastic process: the information drift. arXiv preprint arXiv:1902.06780, 2019.

Leo Neufcourt. Expansion of a filtration with a stochastic process: a high frequency trading perspective. PhD thesis, Columbia University, 2017.

Martin D. Gould, Mason A. Porter, Stacy Williams, Mark McDonald, Daniel J. Fenn, and Sam D. Howison. Limit order books. Quantitative Finance, 13(11):1709–1742, 2013.

Maureen O'Hara. High frequency market microstructure. Journal of Financial Economics, 116(2):257–270, 2015.

Maxi San Miguel and Raul Toral. Stochastic Effects in Physical Systems. In Enrique Tirapegui, Servet Martinez, Enrique Tirapegui, Javier MartÃnez, and Rolando Tiemann, editors, Instabilities and Nonequilibrium Structures VI, volume 5, pages 35–127. Springer Netherlands, Dordrecht, 2000.

Michael Benzaquen and Jean-Philippe Bouchaud. A fractional reaction–diffusion description of supply and demand. The European Physical Journal B, 91(2):23, 2018.

Michael Ostrovsky. Information aggregation in dynamic markets with strategic traders. Econometrica, 80(6):2595–2647, 2012.

Michael Wilkinson. Perturbation Theory for a Stochastic Process with Ornstein-Uhlenbeck Noise. Journal of Statistical Physics, 139(2):345–353, April 2010.

O. E. Barndorff-Nielsen, P. Reinhard Hansen, A. Lunde, and N. Shephard. Realized kernels in practice: trades and quotes. Econometrics Journal, 12(3):C1–C32, November 2009.

P. Bak, M. Paczuski, and M. Shubik. Price variations in a stock market with many agents. Physica A: Statistical Mechanics and its Applications, 246(3-4):430–453, December 1997.

Paul Jusselin, Thibaut Mastrolia, and Mathieu Rosenbaum. Optimal auction duration: A price formation viewpoint. Available at SSRN 3399151, 2019.

Peter J. Rousseeuw and Christophe Croux. Alternatives to the median absolute deviation. Journal of the American Statistical association, 88(424):1273–1283, 1993.

Peter Lakner, Josh Reed, and Florian Simatos. Scaling limit of a limit order book model via the regenerative characterization of levy trees. Stochastic Systems, 7(2):342–373, 2017.

Peter R Hansen and Asger Lunde. Realized Variance and Market Microstructure Noise. Journal of Business & Economic Statistics, 24(2):127–161, April 2006.

Pierre Collin-Dufresne and Vyacheslav Fos. Do Prices Reveal the Presence of Informed Trading?: Do Prices Reveal the Presence of Informed Trading? The Journal of Finance, 70(4):1555–1582, August 2015. ISSN 00221082.

Pierre Collin-Dufresne and Vyacheslav Fos. Insider trading, stochastic liquidity, and equilibrium prices. Econometrica, 84(4):1441–1475, 2016.

Ricardo A Maronna and Ruben H Zamar. Robust Estimates of Location and Dispersion for High-Dimensional Datasets. Technometrics, 44(4):307–317, November 2002.

Richard Haberman. Elementary applied partial differential equations: with Fourier series and boundary value problems. Prentice-Hall, Englewood Ciffs, N.J, 2nd ed edition,1987.

Robert F. Engle and Jeffrey R. Russell. Forecasting transaction rates: The autoregressive conditional duration model. Working Paper 4966, NBER, December 1994.

Roger M Cooke, Harry Joe, and Bo Chang. Vine copula regression for observational studies. AStA Advances in Statistical Analysis, pages 1–27, 2019.

S. Denisov, V. Zaburdaev, and P. Hanggi. Levy walks with velocity fluctuations. Physical Review E, 85(3), March 2012.

Sanford J. Grossman and Merton H. Miller. Liquidity and Market Structure. The Journal of Finance, 43(3):617–633, July 1988. ISSN 00221082.

Shmuel Baruch and Lawrence R Glosten. Tail expectation and imperfect competition in limit order book markets. Journal of Economic Theory, 183:661–697, 2019.

Tarun Chordia, Richard Roll, and Avanidhar Subrahmanyam. Order imbalance, liquidity, and market returns. Journal of Financial Economics, 65(1):111–130, July 2002. ISSN:0304405X.

Thierry Foucault, Ohad Kadan, and Eugene Kandel. Limit order book as a market for liquidity. The review of financial studies, 18(4):1171–1217, 2005.

Thomas E. Copeland and Dan Galai. Information Effects on the Bid-Ask Spread. The Journal of Finance, 38(5):1457, December 1983. ISSN 00221082.

Tilmann Gneiting and Adrian E Raftery. Strictly proper scoring rules, prediction, and estimation. Journal of the American statistical Association, 102(477):359–378, 2007.

V. Zaburdaev, S. Denisov, and J. Klafter. Levy walks. Reviews of Modern Physics, 87(2): 483–530, June 2015.

Vasily Zaburdaev, Michael Schmiedeberg, and Holger Stark. Random walks with random velocities. Physical Review E, 78(1), July 2008.

Vladimir V Palyulina, Aleksei V Chechkinb, and Ralf Metzlera. Levy flights do not always optimize random blind search for sparse targets. PNAS, 111(8):2931–2936, 2014.

Wolfgang Pesendorfer and Jeroen M Swinkels. Efficiency and information aggregation in auctions. American Economic Review, 90(3):499–525, 2000.

Wolfgang Pesendorfer and Jeroen M Swinkels. The loser's curse and information aggregation in common value auctions. Econometrica: Journal of the Econometric Society, pages 1247–1281, 1997.

Yacine Ait-Sahalia and Jialin Yu. High frequency market microstructure noise estimates and liquidity measures. The Annals of Applied Statistics, 3(1):422–457, March 2009.

Younes Kchia and Philip Protter. Progressive filtration expansions via a process, with applications to insider trading. International Journal of Theoretical and Applied Finance, 18(04):1550027, 2015

CURRENT STATUS OF PUBLICATIONS IN PEER REVIEWED JOURNALS

There are four papers emerging from this dissertation. The current status is listed below:

1. Pani, Sudhanshu, 2020, Intraday Trading in Order Driven Markets., (Unpublished, Presented at World Finance Conference, Aug 2021)
2. Pani Sudhanshu, 2020, A theory of 'Auction as a Search' in speculative markets, *International J. of Financial Markets and Derivatives*, (Published, ABDC "C")
3. Pani, Sudhanshu , 2020, Liquidity in High Resolution in Order Driven Markets, *International J. of Financial Markets and Derivatives* , (Published, ABDC "C")
4. Chakravarty, R.R., Pani, Sudhanshu. A Data Paradigm to Operationalise Expanded Filtration: Realized Volatilities and Kernels from Non-Synchronous NASDAQ Quotes and Trades. *Journal of Quantitative Economics,* 19, 617–652 (2021). (Published, ABDC "B", IGIDR-IIM Udaipur Conference, February 2020 and at World Finance Conference, Aug 2021)